Investigation of the Clostridium difficile Sortase by Gene Knockout, X-ray Crystallography and Biochemical Characterisation

Chambers, Christopher

Award date:
2014

Awarding institution:
University of Bath

Investigation of the *Clostridium difficile* Sortase by Gene Knockout, X-ray Crystallography and Biochemical Characterisation

Christopher James Chambers

A Thesis Submitted for the Degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

December 2013

Abstract

The opportunistic pathogen *Clostridium difficile* is the most common cause of antibiotic-associated diarrhoea, with severity of disease ranging from mild diarrhoea to fulminant pseudomembranous colitis and death. It poses a major burden on healthcare providers, costing millions of pounds each year due to ward closures, isolation measures and prolonged illness. The current antibiotic therapy for *C. difficile* infection is effective, but high rates of relapse lead to ongoing misery for patients and spiralling costs for healthcare providers. Novel therapeutics for *C. difficile* are therefore desperately sought, and the antibiotic-induced nature of the disease has led to interest in development of non-antibiotic therapies.

Sortase enzymes are responsible for covalent anchoring of specific proteins to the peptidoglycan of the cell wall of gram-positive bacteria. Following the discovery that the Sortase A enzyme of *Staphylococcus aureus* is essential for pathogenesis, sortase inhibitors are under investigation as novel therapeutics. Being ubiquitous in gram-positive bacteria, it is likely that other gram-positive pathogens require sortase enzymes for their pathogenesis and may be targets for development of sortase inhibitors.

This work describes a characterisation of the sortase enzyme of *C. difficile*. To provide evidence for a role of the sortase in the cell wall biogenesis, a *C. difficile* sortase knockout strain was constructed by intron mutagenesis. Characterisation of this mutant led to the discovery that the putative adhesin CD0386 is anchored to the peptidoglycan of *C. difficile* by the sortase SrtB. To provide structural insight into the catalytic mechanism of the *C. difficile* sortase, an active site mutant was crystallised and its structure solved to 2.55Å by X-ray diffraction. The wall-linked protein CD0386 was also crystallised and subject to successful test diffraction. Analyses of SrtB reaction products by chromatography and mass spectroscopy indicate that the enzyme cleaves an SPKTG peptide motif and catalyses a transpeptidation reaction with a component of the *C. difficile* peptidoglycan.

Acknowledgements

I wish to offer my thanks first and foremost to my supervisors Dr. Cliff Shone (Public Health England (PHE), Porton Down) and Prof. Ravi Acharya (University of Bath) for their guidance, motivation and expertise. I would also like to extend my thanks to Dr. April Roberts (PHE) for her knowledge in all matters *Clostridium difficile* and general assistance in administration and execution of the project. Thank you to Public Health England for funding and to the Diamond Light Source (Oxon, UK) for access to synchrotron facilities.

Thank you to all members of the Toxins Group at Public Health England and Lab 0.34 at the University of Bath, for always making me feel at home wherever I happened to be working. Special thanks are owed to Dr Jon Kirby (PHE) for his assistance with the Clostron system, to Dr Nethaji Thiyagarajan (Bath) for his assistance with computing, and to Helen and Dan Ahern for being brilliant next-door neighbours.

Last but of course not least, thank you to my parents for supporting me through 22 years of education (and everything else in between).

Declaration of Work Performed in Conjuction with Others

Mass spectroscopy was performed by Dr Anneke Lubben of the Chemical Characterisation and Analysis Facility, University of Bath. Collection of X-ray diffraction data at the Diamond Light Source was kindly performed by Dr Abigail Davies and Miss Charlotte Harrison.

In vivo procedures were performed by staff of Public Health England, Porton Down, under a UK Home Office-approved Project Licence and in full compliance with the UK Animals (Scientific Procedures) Act 1986.

Table of Contents

Chapter 1 –Introduction to *Clostridium difficile*

Chapter 2 - Introduction to the Sortase Enzyme Family

Chapter 3 - General Methods

Chapter 4 – Purification of Recombinant Proteins

Chapter 5 - Characterisation of a *C. difficile* Sortase Knockout

Chapter 8 – Biochemical Characterisation of the *C. difficile* SrtB Reaction

Chapter 9 – General Discussion

Appendix A – Nucleotide and Amino Acid Sequences

List of Figures

Chapter 5 – Characterisation of a *C. difficile* Sortase Knockout

Chapter 6 - Structural Characterisation of the *Clostridium difficile* Sortase Enzyme

List of Tables

List of Abbreviations

A. oris	*Actinomyces oris*
ADP	Adenosine Triphosphate
B. anthracis	*Bacillus anthracis*
B. cereus	*Bacillus cereus*
BCA	Bicinchoninic acid
BHI	Brain Heart Infusion
C. bifermentans	*Clostridium bifermentans*
C. difficile	*Clostridium difficile*
C. diphtheriae	*Corynebacterium diptheriae*
C. perfringens	*Clostridium perfringens*
CCD	Charge Coupled Device
CDI	*Clostridium difficile* infection
CDT	*Clostridium difficile* transferase
CWSS	Cell Wall Sorting Signal
DAP	Diaminopimelic acid
DLS	Diamond Light Source
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ELISA	Enzyme Linked Immunosorbent Assay
ESI-MS	Electrospray Ionisation Mass Spectroscopy
FCS	Foetal Calf Serum
FRET	Förster Resonance Energy Transfer
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HPLC	High Pressure Liquid Chromatography

IEP	Intron Encoded protein
IPTG	Isopryl-1-thio-β-galactopyranoside
L. monocytogenes	*Listeria monocytogenes*
LB	Lysogeny Broth
LC-MS	Liquid Chromatography Mass Spectroscopy
LLG	Log Likelihood Gain
M. tuberculosis	*Mycobacterium tuberculosis*
MAD	Multiple Wavelength Anomalous Dispersion
mDAP	meso-Diaminopimelic Acid
MIC	Minimum Inhibitory Concentration
MIR	Multiple Isomorphous Replacement
NCS	Non Crystallographic Symmetry
NEAT	Near Iron Transporter
NMR	Nuclear Magnetic Resonance
ORF	Open Reading Frame
PAK	Packing Clashes
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PMC	Psuedomembranous Colitis
PPI	Proton pump inhibitor
RAM	Retrotransposition activatible marker
RFZ	Rotation Function Z-Score
RMSD	Root mean standard deviation
RNA	Ribonucleic Acid
RTPCR	Reverse Transcriptase Polymerase Chain Reaxction
S. agalactiae	*Streptococcus agalactiae*
S. aureus	*Staphylococcus aureus*
S. gordonii	*Streptococcus gordonii*
S. mutans	*Streptococcus mutans*
S. pyogenes	*Streptococcus pyogenes*
SAD	Single Wavelength Anomalous Dispersion

SCOP	Structural Classification of Proteins
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SLP	S Layer Protein
Srt[A-D]	Sortase Class [A-D]
TB	Terrific Broth
TCEP	Tris-(2-carboxyethyl)phosphine
TFZ	Translation Function Z-Score
TMB	3,3',5,5'-Tetramethylbenzidine
Tris	Tris(hydroxymethyl)aminomethane

Chapter 1 – Introduction to *Clostridium difficile*

1.1.1 Introduction to *Clostridium difficile*

Clostridium difficile is an obligately anaerobic Gram-positive bacterium that exhibits a rod-shaped vegetative cell morphology and forms cylindrical endospores (Figure 1.1.1). It was first isolated from the faeces of healthy infants and originally named *Bacillus difficilus* due to difficulties that early investigators encountered in its isolation (Hall and O'Toole, 1935). The toxigenic nature of *C. difficile* was rapidly identified in initial studies - in guinea pigs, culture filtrates caused subcutaneous oedema, convulsions and death. Despite its apparent lethality, the newly discovered organism was isolated from healthy subjects and did not appear to be a human pathogen, so with the exception of sporadic reports of its isolation (Smith and King, 1962), the bacterium received little attention from the scientific community for around 40 years.

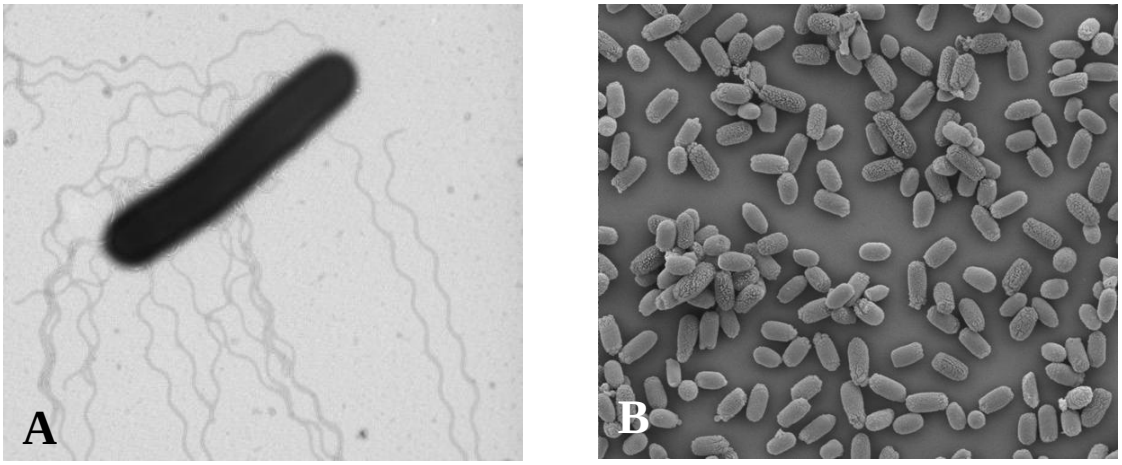


Figure 1.1.1 – *Clostridium difficile.* (A) Transmission electron microscopy of *C. difficile* reveals a rod-like morphology with peritrichous flagella (B) Scanning electron micrograph of highly purified *C. difficile* endospores *(Images courtesy of Electron Microscopy Unit, Public Health England (PHE) Porton).*

This period of relative obscurity came to an end in the mid-1970s, amidst anecdotal reports of pseudomembranous colitis (PMC) in patients undergoing therapy with clindamycin, a recently introduced lincosamide antibiotic (Fenton *et al.*, 1974; Miller and Wheeler, 1975). These concerns led to Upjohn, the manufacturers of clindamycin, to fund a prospective study which found that of 200 patients undergoing clindamycin therapy, 21% suffered diarrhoea and 10% suffered PMC (Tedesco *et al.*, 1974).

After establishing that clindamycin induced colitis was transferable in hamsters by passage of cecal contents even after removal of clindamycin, it was found that the agent causing disease could not

be filtered from the cecal contents with a 0.02μm filter but was successfully removed using a filter with a molecular weight limit of 10kDa (Bartlett *et al.*, 1977). In addition, it was found that the cytopathic effect could be blocked by antiserum against *Clostridium perfringens* toxin. Concurrently, other groups examined the cytopathic effect of faecal extracts of clindamycin-associated colitis patients, again implicating a heat labile toxin of >6 kDa which could be neutralised *in vitro* with antibodies against *Clostridium sordellii* toxin (Rifkin *et al.*, 1977). Culture *of Clostridium spp*. from PMC patients resulted in isolation of several toxigenic *Clostridium* species, but *C. difficile* was the only species which produced the characteristic cytopathic effects *in vitro* (George *et al.*, 1978). Thus, by the end of the 1970s, there was good evidence that antibiotic-associated PMC was caused by a toxin or toxins secreted by *C. difficile*.

1.1.2 Asymptomatic Carriage of *C. difficile*

Carriage of *C. difficile* in infants and neonates is common, with estimates from single time point screening ranging from 2% to 84% (Larson *et al.*, 1982; Matsuki *et al.*, 2005). In a longitudinal study of 10 healthy infants, all tested positive at some point during the one year study period, suggesting that colonisation with *C. difficile* may be a universal albeit transient phenomenon (Rousseau *et al.*, 2012). Asymptomaticity is a notable characteristic of colonisation in early life despite detectable levels of toxin in the faeces (Stark *et al.*, 1982). Carriage rate declines after the first year of life, but a small proportion (<10%) of adults are colonised (Loo *et al.*, 2011; Miyajima *et al.*, 2011), with much higher rates observed in certain populations such as those residing in long term elderly care facilities (Riggs *et al.*, 2007).

1.1.3 *C. difficile* disease and risk factors

As an opportunistic pathogen, *C. difficile* can cause a wide spectrum of disease, ranging from mild diarrhoea and abdominal pain to pseudomembranous colitis (See Figure 1.1.2), bowel perforation, sepsis and toxic megacolon (Rupnik *et al.*, 2009). *Clostridium difficile* Infection (CDI) is frequently fatal, with typically reported 30-day fatality rates in the range of 10-35% (Hota *et al.*, 2012; Kotila *et al.*, 2011). Definitive diagnosis of CDI is made by the detection of the presence of the classical *C. difficile* toxins TcdA and TcdB in the faeces. Historically this was performed by cytotoxicity testing, but is now routinely determined by imunosorbent assay (Kvach *et al.*, 2010) or by detection of the *tcdA* and *tcdB* genes by nucleic acid amplification (De Jong *et al.*, 2012).

The transition of *C. difficile* from harmless commensal organism to lethal pathogen is thought to be due to changes in the normal colonic flora during antimicrobial chemotherapy, providing an ecological niche which is then exploited by *C. difficile*. Although clindamycin was the antibiotic originally implicated in CDI, many other antibiotics have since been identified as risk factors. In a

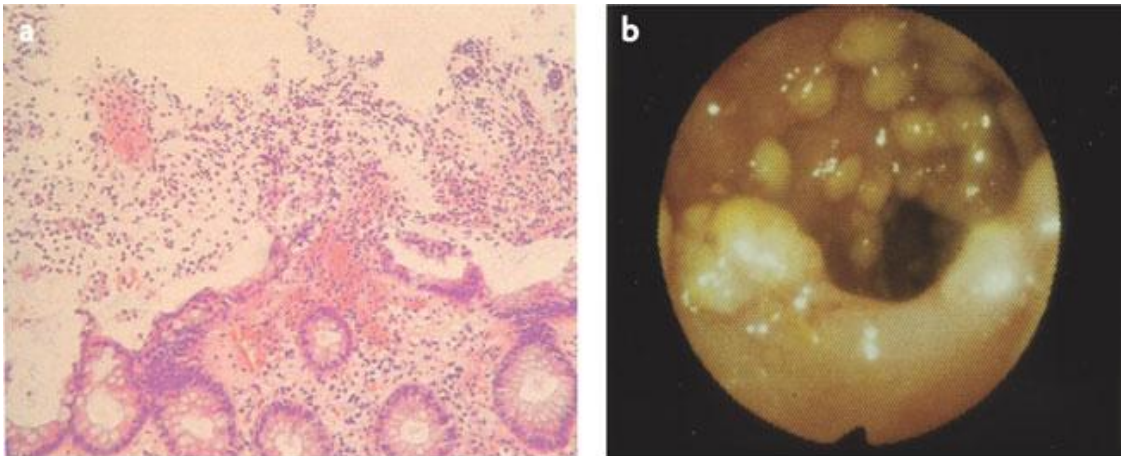


Figure 1.1.2 - Gross and Microscopic Pathology of *C. difficile* Infection. A) Disruption of the epithelial layer resulting in 'volcano-like' microscopic lesions. B) Pseudomembranous lesions in the ascending colon visualised by colonoscopy (Rupnik *et al.*, 2009).

1998 meta-analysis of 49 studies, the third generation cephalosporins Cefotaxime, Ceftazidim, Cefoxitin and Cefuroxime all scored odds ratios for development of CDI higher than that of Clindamycin (Bignardi, 1998). An added difficulty with the use of cephalosporins is that in addition to disruption of the host gut flora, they have little effect on *C. difficile* itself, which appears to be universally resistant (Gerding, 2004). More recently, the use of fluoroquinolones, which was once considered low risk, has emerged as a risk factor of even greater significance than cephalosporins (McCusker *et al.*, 2003). This change in epidemiology has followed the emergence of strains with high levels of resistance to fluoroquinolones due to a single base substitution in the *gyrA* gene encoding DNA Topoisomerase IV, the target of quinolone antibiotics (Drudy *et al.*, 2007b)

The effects of clindamycin treatment on the intestinal microbiota in mice and hamsters have been studied in depth with the use of metagenomic studies of 16S ribosomal RNA (16S rRNA) (Buffie *et al.*, 2012; Peterfreund *et al.*, 2012). The overall 16S rRNA copy number observed in the ileum or cecum did not change after clindamycin administration in mice. Rather, there was a reduction in the number of *unique* 16S rRNA sequences, indicating a reduction in microbial diversity rather than a reduction in population. In particular, clindamycin treatment resulted in an almost complete obliteration of *Bacteroides spp*, *Ruminococcus spp,* and members of the *Lachnospiraceae* family. Expansions of population were observed in other bacterial species, reflected in the observation that overall 16S rRNA copy number was similar before and after clindamycin administration. Hence in mice, clindamycin causes changes in the intestinal microbiota rather than a reduction in numbers. Experiments in humans have demonstrated that clindamycin induces remarkably long-lasting changes in intestinal microbiota, with a reduction in *Bacteriodes* diversity even 24 months after treatment (Jernberg *et al.*, 2007)

Association of disease with advanced age is almost universal in epidemiological studies of CDI, with estimates of increased risk ranging from 2-4% per year after the age of 18 (Loo *et al.*, 2011; Starr *et al.*, 2003). The identification of gastric acid suppression by H_2 receptor agonists or Proton Pump Inhibitors (PPIs) as a risk factor for CDI (Cunningham *et al.*, 2003; Walker *et al.*, 1993) has proven controversial, as some studies have found no statistically significant association (Pepin *et al.*, 2005a). Meta analyses have, however, demonstrated that PPIs are likely to be a risk factor, with odds ratios ranging from 1.5 to 2.1 (Janarthanan *et al.*, 2012; Kwok *et al.*, 2012). *C. difficile* spores are capable of outgrowth in the gastric contents of patients undergoing therapy with PPIs, but only when supplemented with taurocholic acid and amino acids, so this is unlikely to occur *in vivo* (Nerandzic *et al.*, 2009). *In vitro*, vegetative *C. difficile* cells are able to survive longer in the stomach contents of patients undergoing therapy with PPIs than those who are not – in some patients undergoing PPI treatment, the pH of the stomach contents was as high as pH7 and survival of *C. difficile* in these conditions is not significantly different to survival in pH7 phosphate buffered saline (Jump *et al.*, 2007). Enhanced survival of vegetative cells in the stomach may therefore account for the increased risk of CDI due to acid suppression.

1.1.4 Emergence of Hypervirulent/Epidemic *C. difficile*

In the early 2000s, several retrospective studies uncovered significant changes in *C. difficile* epidemiology and worsening patient outcomes during the late 1990s (Dallal *et al.*, 2002; Morris *et al.*, 2002; Pepin *et al.*, 2004). For example, in the study by Pepin *et al*. in Quebec, the proportion of patients who died within 30 days of diagnosis was 4.7% in 1991-92 but had risen to 13.8% in 2003. A detailed study of the Quebec outbreak identified the predominant strain as an 027 ribotype, and it was suggested that this strain may be a novel 'hypervirulent' type (Pepin *et al.*, 2005b). Similar strains of ribotype 027 were isolated from patients in two prolonged and severe outbreaks of *C. difficile* at Stoke Mandeville hospital in the UK, where 498 people were infected and 117 died between April 2003 and March 2006 (Healthcare Commission, 2006). Surveillance in 2005 revealed that 25% of cultures from symptomatic patients in England and Wales were ribotype 027, a figure which rose to 41% in 2007 (Brazier *et al.*, 2008). As shown in Figure 1.1.3, a considerable increase in mortality from CDI was observed during this period.

Although many authors frequently describe ribotype 027 strains as “hypervirulent” and state a link between ribotype 027 and increased severity of disease, not all of the evidence supports this assertion. For example, a retrospective case-matched study in the UK comparing 123 patients with 027 or non-027 strains failed to find any association between 027 strains and increased severity (Morgan *et al.*, 2008). Similar studies in the USA (Walk *et al.*, 2012) and France (Sirard *et al.*, 2011) failed to find any association between disease severity and ribotype. Conversely, a study of

863 patients from across the Netherlands found that when compared with non-027 strains, ribotype 027 were more likely to cause severe diarrhoea (22.4% vs. 13.9%) and relapses were more common (20.6% vs. 14.9%) (Goorhuis *et al.*, 2007). Given the undoubted success of ribotype 027 in causing epidemics, many authors prefer to describe it as an 'epidemic' rather than 'hypervirulent' strain.

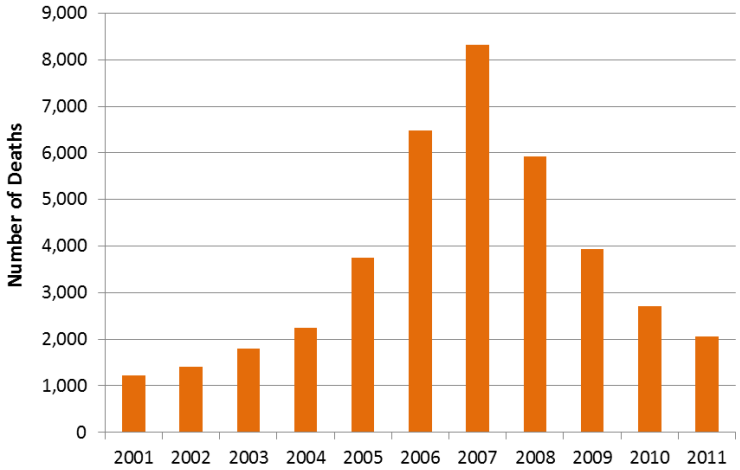


Figure 1.1.3 – Deaths Attributed to *C. difficile* in England and Wales in 2001-2011. A significant peak in mortality was observed 2005-2010, concomitant with the dominance of epidemic 027 strains (Statistics collected by the Office of National Statistics).

Regardless of their *in vivo* virulence, 027 ribotype strains do exhibit *in vitro* characteristics which could account for hypervirulence. Most notable among these is increased toxin production, with epidemic 027 strains producing up to 16 times more toxin A and 23 times more toxin B than toxinotype 0 strains (Warny *et al.*, 2005), a phenomenon discussed in Section 1.3.3.

Several investigators have reported an increased rate of sporulation in epidemic 027 strains compared to non-027 (Akerlund *et al.*, 2008; Merrigan *et al.*, 2010). However, testing of a much larger sample of isolates found substantial variation in sporulation rate but no statistically significant association of high sporulation with epidemic 027 strains. Statistically significant diversity in sporulation rate was observed even within the group of epidemic 027 strains isolated within the UK (Burns *et al.*, 2011, 2010) These results may indicate that previous findings of high sporulation in epidemic 027 strains may have been due to unintentional selection of highly sporulating strains.

1.2 – The *C. difficile* Cell Surface

1.2.1 Overview of the *C. difficile* Cell Surface

The surface of *C. difficile* is typical of a Gram-positive bacterium, composed of a lipid membrane, a substantial layer of peptidoglycan and an additional exterior proteinacous layer named the Surface or 'S' Layer (Figure 1.2.1).

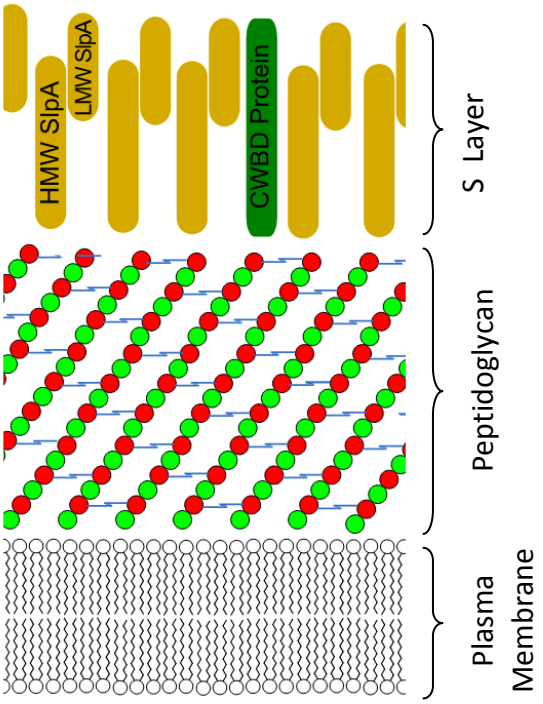


Figure 1.2.1 – The *C. difficile* Cell Envelope. In common with many Gram Positive bacteria, *C. difficile* possesses an additional proteinaceous layer exterior to the peptidoglycan, named the surface layer (S layer), composed of a self-assembling paracrystalline array of high (HMW) and low (LMW) molecular weight SlpA interspersed with Cell Wall Binding Domain (CWBD) proteins.

1.2.2 The Peptidoglycan of *C. difficile*

A comprehensive analysis of the *C. difficile* peptidoglycan by mass spectroscopy of muramidase-digested peptidoglycan fragments (Peltier *et al.*, 2011), indicates the sequence of the tetrapeptide is L-Ala-D-Glu-A_2pm-D-Ala, where A_2pm is diaminopimelic acid (DAP). Of these tetrapeptides, 41% are 3-3 linked (DAP to DAP, L,D transpeptidated) while 15% are 4-3 linked (D-ala to DAP D,D-transpeptidated). Diaminopimelic acid has two stereogenic centres and hence can exist in three stereoisomers, DD-, LL- and *meso*-. The chirality of the DAP in *C. difficile* peptidoglycan has not yet been determined empirically, but Peltier *et al* (2011) speculate that it likely to be *meso*-stereoisomer (mDap), based on the evidence that of the 32 *Clostridium* species previously examined, 28 contained the *meso*- stereoisomer (Schleifer and Kandler, 1972) including *C. bifermentans* and *C. perfringens*, which are considered the closest genetic relatives of *C. difficile* (Collins *et al.*, 1994).

A high level of N-deacetylation was observed in *C. difficile* peptidoglycan, specifically of N-acetylglucosamine residues, with 93% of these bearing no acetyl group. Modification of sugars in this manner confers resistance to muramidases such as lysozyme and is generally considered a virulence factor (Boneca *et al.*, 2007)

1.2.3 The S-Layer of *C. difficile*

In most species in which an S-layer has been characterised, the S-layer is composed of a single predominant protein. In some species, the S-layer is composed of two proteins but generally these proteins are encoded by separate genes. *Bacillus anthracis* for example, possesses two major SLPs named EA1 and Sap (Fouet *et al.*, 1999). *C. difficile* is unusual in that it possesses two distinct S-layer components, produced by the cleavage of a single precursor SlpA by the cysteine protease Cwp84 (Calabi *et al.*, 2001; Kirby *et al.*, 2009). The SlpA cleavage products re-associate non-covalently via the N terminus of the high molecular weight (HMW) fragment and the C terminus of the low molecular weight (LMW) fragment (Fagan *et al.*, 2009), forming a crystalline array of equimolar quantities of the HWM and LWM proteins (Masuda *et al.*, 1989). The HMW SlpA contains a predicted cell wall binding motif (Pfam motif PF04122), and it is therefore thought that the HMW component binds to the bacterial cell wall while the LMW SlpA forms the exterior of the array. This is confirmed by the observations that LMW SlpA is the immunodominant antigen recognised by convalescent sera so likely to be presented at the cell surface (Pantosti *et al.*, 1989)

Recombinant *C. difficile* SLPs bind to cultured human epithelial (HEP-2) cells, the epithelium and lamina propria of human colon tissues and several purified extracellular matrix components (Calabi *et al.*, 2002). In addition, incubation of *C. difficile* cells with anti-HMW SLP antibodies results in a 20-30% reduction in binding to HEP-2 cells. While this may indicate that the HMW is a physiologically relevant adhesin, the results must be interpreted with care as the surface of *C. difficile* is covered almost entirely in the HMW SLP and therefore binding of antibodies to HMW SLP may result in steric hindrance of other adhesins.

1.2.4 The C. difficile Cell Wall Binding Domain (CWBD) Proteins

The *C. difficile* genome encodes 28 proteins containing repeats of the Pfam motif PF04122, described as the cell wall binding domain (CWBD), a domain which is characteristic of the S-layer protein SlpA and is responsible for its surface localisation.

Cwp84, the cysteine protease which cleaves SlpA, contains three repeats of the PF04122 motif and is expressed as a proenzyme and proteolytically processed to its active form (Chapetón Montes *et al.*, 2011). It was originally thought that this cleavage may be an autocatalyic process similar to

other papain family enzymes, but mutation of the Cwp84 active site cysteine to alanine does not have any effect on its processing (de la Riva *et al.*, 2011). Insertional inactivation of the closely homologous *cwp19* gene, encoding another CWBD protein, does however result in a partial loss of Cwp84 processing. The residual processing of Cwp84 indicates that while Cwp19 may have a role in Cwp84 cleavage, either another protein must be involved or Cwp84 possesses catalytic activity independent of its putative papain-like active site. Cwp19 also exhibits some proteolytic activity towards SlpA, but this is insufficient to complement a cwp84 deficient mutant (de la Riva *et al.*, 2011)

Several other CWBD proteins have been investigated experimentally and exhibit properties which may be linked to virulence. For example, the CWBD protein Cwp66 has been implicated in adhesion to mammalian cells (Waligora *et al.*, 2001). The CWBD protein CwpV exhibits phase variable expression due to a genetic inversion (Emerson *et al.*, 2009) and may therefore be involved in immune evasion, and although its function is currently unclear, it appears to promote aggregation in some *C. difficile* strains (Reynolds *et al.*, 2011).

1.3 – The *C. difficile* Toxins

1.3.1 The Classical Toxins TcdA and TcdB

The genes encoding the classical *C. difficile* toxins TcdA and TcdB are located in a 19kb element termed the pathogenicity locus (PaLoc). Both toxins have a domain structure comprising an N-terminal glucosyltransferase domain, an autocatalytic cysteine protease domain, a translocation domain and a C-terminal receptor binding domain (Egerer *et al.*, 2007).

The binding domain of TcdA binds to gp96, a glycoprotein present on the surface of human colonocytes (Na *et al.*, 2008), likely via a carbohydrate moiety (Teneberg *et al.*, 1996). The receptor for TcdB remains unknown. Following cellular entry of TcdA and TcdB via clathrin mediated endocytosis (Papatheodorou *et al.*, 2010) and in the low pH of the acidified endosome, structural changes within the translocation domain lead to the formation of channels in the membrane through which the glucosyltransferase and cysteine protease domains can pass (Barth *et al.*, 2001; Giesemann *et al.*, 2006). The cysteine protease domain is activated by binding of cytoplasmic inositol-6-phosphate (Egerer *et al.*, 2009) and cleaves the toxin to release the glucosyltransferase domain into the cytoplasm while the binding and translocation domains remain membrane associated (Pfeifer *et al.*, 2003)

Once in the cytoplasm, the glucosyltransferase domains of toxins TcdA and TcdB selectively O-glucosylate threonine 37 of RhoA and the equivalent positions of the Rho family proteins Rac and

Cdc42, utilising UDP glucose as a cofactor. (Just *et al.*, 1995a, 1995b). The resulting dysfunction in the cytoskeleton leads to cell death and disruption of the epithelium. Despite the structural homology of Toxins A and B, there are apparent differences in their activity and relative clinical importance, with disagreement between *in vitro*, *in vivo* and clinical data. The most definitive experiments investigating the relative effects of TcdA and TcdB have been enabled by the systematic generation of isogenic mutants which are either *tcdA+ tcdB-* or *tcdA- tcdB+*. In a hamster model, it was found that *tcdA- tcdB+* mutants exhibited similar virulence to the *tcdA+ tcdB+* wild type, whereas *tcdA+ tcdB-* mutants were significantly attenuated, with greater levels of survival and significantly longer time from colonisation to death in non survivors (Lyras *et al.*, 2009). That TcdB is not only essential but also sufficient for disease is supported by the observation that *tcdA- tcdB+* strains are frequently isolated from CDI patients including in outbreaks (Alfa *et al.*, 2000; Drudy *et al.*, 2007a). However, an antibody protection study found that upon challenge with a *tcdA+ tcdB+* strain, systemically delivered anti-TcdB antibodies did not offer any protection as would be expected if TcdB was essential for disease (Roberts *et al.*, 2012). In conclusion, there remains conflicting evidence on the relative contributions to pathogenesis of the two classical toxins despite several investigations.

1.3.2 The *C. difficile* Binary Toxin, *Clostridium difficile* Transferase

The third *C. difficile* toxin, named *Clostridium difficile* Transferase (CDT) (Popoff *et al.*, 1988) is a binary toxin encoded by the genes *cdtA* and *cdtB*, located outside of the PaLoc. CDT is a member of the family of binary actin ADP-ribosylating toxins including the C3 toxin from *Clostridium botulinum*, the iota toxin of *Clostridium perfringens* and the vegetative insecticide protein of *Bacillus cereus*. The proportion of strains which possess the *cdtA* and *cdtB* genes is relatively low - retrospective surveys of strain collections have revealed *cdtAB* gene carriage of between 6.4% and 5.8% (Geric, 2004; Stubbs *et al.*, 2000). Epidemic 027 strains are typically *cdtAB*$^{+}$. (Stabler *et al.*, 2009)

Assuming structural and functional homology with the other members of the family, CDTa is monomeric while CDTb forms heptamers on proteolytic activation (Barth *et al.*, 2000). The holotoxin is endocytosed following binding of CDTb to the lipolysis stimulated lipoprotein receptor (Papatheodorou *et al.*, 2011), following which acidification of the endosome induces conformational changes within the CDTb heptamer promoting membrane insertion and pore formation (Barth *et al.*, 2000). The CDTa component then translocates through the pore and into the cytoplasm, where it ADP-ribosylates actin (Gülke *et al.*, 2001) likely to be at arginine residue 177 as is the case in *C. botulinum* C2 toxin *and C. perfrignens* iota toxin (Vandekerckhove *et al.*, 1988). ADP ribosylation of this residue prevents incorporation of actin into actin filaments and

caps the barbed ends of actin filaments (Wegner and Aktories, 1988) leading to a shift in actin dynamics favouring depolymerisation. Application of CDT to cultured cells results in cell rounding, depolymerisation of actin and cell death (Perelle *et al.*, 1997). In CaCo-2 colon carcinoma cells, disorder in microtubules has been observed, resulting in protrusions from the cell which increase the adherence of *C. difficile in vitro* (Schwan *et al.*, 2009).

The clinical relevance of CDT is still unclear. A small (n=26) case-control study (Barbut *et al.*, 2005) found that patients infected with *cdtAB+ C. difficile* strains suffered more severe disease than those with *cdtAB* $^{-}$ strains. A larger (n=131) study (Barbut *et al.*, 2007) also found association with greater severity and another (Goldenberg 2011, n = 207) found a significantly higher all-cause mortality at 30 days (31% vs 14%). However, none of these studies had sufficient sample size to conduct multivariate analysis – this is vital for a proper analysis because CDT carriage is clonal, requiring multivariate analysis to determine its independence from other strain dependent factors.

1.3.3 Regulation of Toxin Production

The PaLoc gene *tcdR* encodes a 22kDa lysine rich protein with homology to putative toxin regulators in *C. botulinum* and *C. perfringens* (Moncrief *et al.*, 1997). Gel shift and *in vitro* transcription analysis revealed that TcdR does not bind directly to the PaLoc promoters , but is an alternate sigma factor, allowing the RNA polymerase complex to bind to and initiate transcription from the *tcdA* and *tcdB* promoters (Moncrief *et al.*, 1997). As a putative anti-sigma-factor, TcdC was suggested to be a negative regulator, antagonistic to TcdR. Accordingly, in *C. perfringens*, transcription of a fusion protein under the control of the *C. difficile tcdA* promoter is inhibited by the co-expression of TcdC. In the same work, *in vitro* expression experiments with *E. coli* RNA polymerase (RNAP) complexes confirmed that TcdC disrupts the interaction between the RNAP and TcdR, confirming its activity as an anti sigma factor (Matamouros *et al.*, 2007).

Early genomic characterisations of ribotype 027 strains revealed an 18bp deletion at position 330 of *tcdC* and a single base pair deletion at position 117, resulting in a frameshift and hence a truncation of TcdC (MacCannell *et al.*, 2006). Based on the evidence of TcdC anti-sigma activity, it was therefore suggested that a deregulation of toxin production due to loss of functional TcdC was responsible for increased toxin production of epidemic 027 isolates. To investigate the role of the 18bp deletion in tcdC observed in some strains, Carter and co-workers introduced a full length, functional tcdC from *C. difficile* VPI10463, a known high producer of toxin, into an 027 strain in trans (Carter *et al.*, 2011). This resulted in a significant reduction in toxin titres and a reduction in virulence in a hamster model.

However, the restoration of functional *tcdC* in an 027 ribotype strain by allele exchange has no effect on toxin production (Cartman *et al.*, 2012). Similarly, when *tcdC* is inactivated in a non-027 strain, only a modest increase in *tcdRBEA* transcription is observed, and there is no detectable increase in secreted toxins (Bakker *et al.*, 2012). While these results clearly cast doubt on the role of TcdC as a repressor of toxin production, it is possible that TcdC has a more specialised role in regulation, perhaps regulating toxin production in conditions not adequately replicated in the above experiments.

Global regulation of the toxin production is linked to nutritional stress – transcription of toxin genes is low during exponential growth when nutrients are plentiful or when artificially repressed by glucose *in vitro*, and increases during the stationary phase as nutrients become scarce. This toxigenic response to nutritional stress is co-ordinated by the transcriptional regulator CodY, a transcriptional regulator whose affinity for its cognate promoters is modulated by the availability of branched chain amino acids, a proxy for nutritional sufficiency (Shivers and Sonenshein, 2004). In the absence of CodY, *tcdRBEA* transcription is increased between 51- and 174-fold (Dineen *et al.*, 2010).

1.4 Current and Future Therapy of *C. difficile* disease

1.4.1 Chemotherapeutics

Vancomycin is considered the gold standard treatment *for C. difficile*. It is a glycopeptide produced by the bacterium *Amycolatopsis orientalis* and inhibits bacterial growth by binding to C-terminal D-ala-D-ala residues of peptidoglycan pentapeptides, preventing further transglycosylation or transpeptidation and thus preventing synthesis of functional peptidoglycan (Courvalin, 2006). Due to its negligible absorbance in the grastrointestinal tract, vancomycin concentrations of >2000mg/L are achievable in the colon (Gonzales *et al.*, 2010), compared to a typical *C. difficile* minimum inhibitory concentration (MIC) of 0.5-2mg/L. (Aspevall *et al.*, 2006). There are no reports of *C. difficile* isolates resistant to vancomycin. However, rates of CDI relapse after treatment with vancomycin are high, estimated at 15-25%, with many patients suffering multiple relapses (Louie *et al.*, 2011; Wenisch *et al.*, 1996) or reinfection with another strain (Wilcox *et al.*, 1998). This relapsing effect is generally ascribed to the continuing suppression of colonic flora by vancomycin, preserving the flora-depleted niche that led to *C. difficile* disease in the first instance.

Despite its effectiveness, the use of vancomycin for treatment of CDI must be considered carefully due to its considerable cost and growing concerns about vancomycin resistance in *Enterococcus spp*. and *Staphylococcus spp*. For these reasons metronidazole is generally preferred as a first-line

treatment, with trials indicating that it is similarly effective as vancomycin in mild or moderate CDI (Zar *et al.*, 2007). As with vancomycin, typical MICs for metronidazole are low, in the range of 0.032-1.0 mg/L (Aspevall *et al.*, 2006), although high levels of absorption from the gastrointestinal tract result in much lower (<50mg/g) concentrations in the colon (Bolton and Culshaw, 1986). Rates of relapse are similar to those seen with vancomycin therapy (Zar *et al.*, 2007), and although reduced susceptibility to metronidazole is frequently observed it is not widespread and seems to be restricted to ribotypes 001 and 10 (Baines *et al.*, 2008; Brazier, 2001)

In 2011, the antibiotic fidaxomycin was licensed in Europe for treatment of CDI in light of randomised trials demonstrating non-inferiority to vancomycin (Louie *et al.*, 2011). In fact, relapse rates were lower in patients treated with fidaxomycin (13%) than vancomycin (24%). This protective effect may be due to the fact that Fidaxomycin has been demonstrated to have little effect on *Bacteriodes spp*. species which, as discussed in Section 1.1.3, is one of the groups of bacteria most depleted by antibiotics which predispose patients to CDI (Louie *et al.*, 2012)

1.4.2 Probiotics

Probiotic strains investigated for prevention of CDI include *Lactobacillus rhamnosus*, *Lactobacillus casei*, *Lactobacilli helveticus*, *Bifidobacterium longum*, and formulations often include combinations of multiple strains (Allen *et al.*, 2012; Gao *et al.*, 2010; Rampelli *et al.*, 2013). A recent meta-analysis of 23 trials with 4213 participants found that co-administration of probiotics reduced the rate of CDI of those undergoing antibiotic therapy by 64% (Goldenberg *et al.*, 1996). A potential flaw exists in meta analysis of probiotic trials due to the significant heterogeneity of the compounds tested, but the above study did not find statistical evidence of subgroup effects, suggesting that all treatments were similarly effective.

It is unlikely that the mode of action of these therapies is reconstitution of the gut flora, since the strains of bacteria administered (often from the *Lactococcus spp*) are not closely related to those which are absent from the CDI-predisposed colon (typically *Bacteriodes spp*, see section). A more likely hypothesis is inhibition of pathogen growth by acidic secondary metabolites such as lactate and acetate. This hypothesis is supported by the observation that *in vitro* inhibition of *C. difficile* growth by culture supernatants of 10 commonly used probiotic strains is lost when culture supernatants were buffered at pH7 (Schoster *et al.*, 2013).

1.4.3 Surgery

Where disease progresses to toxic megacolon, multiple organ failure, or fails to respond to chemotherapeutics, colectomy is a last resort treatment. Postoperative mortality is high, with a 30-day mortality of 41% (Bhangu *et al.*, 2012)

1.4.4 Faecal Transplantation

Faecal transplantation, the use of faeces from a healthy person to reconstitute the colonic flora, has been used to successfully treat CDI on a sporadic basis for some time, either via colonoscopy infusion or nasogastric tube (Aas *et al.*, 2003; Rohlke *et al.*, 2010). A recent meta-analysis which examined 273 patients from 11 studies found a rate of clinical resolution of 89% (Kassam *et al.*, 2013), which is comparable with the ~90% cure rate observed for vancomycin (Louie *et al.*, 2011)

In the studies reported above, recipients were encouraged to select donors from their immediate family, with the rationale that because people living together have similar faecal microbiota, it was more likely to be successful. However, Kassam *et al* (2013) found no significant difference in outcome between trials with volunteer donors and those with familial donors. Transmission of pathogens is clearly a concern in faecal transplantation, and although such concerns may limit the implementation of faecal transplant, these problems are likely to be surmountable with standardisation and evaluation of testing procedures as is the case for other common procedures such as blood transfusion.

1.4.5 Active and Passive Immunotherapy

There is good evidence that a *C. difficile* vaccine which induces a humoral response could prevent or improve outcomes of CDI patients. Poor anti-toxin antibody titres are significantly associated with a longer duration of disease and a higher risk of relapse (Warny *et al.*, 1994) and patients with a high level of anti-TcdA IgM on day three of their first episode are less likely to experience a relapse (Kyne 2001). As the target population for a *C. difficile* vaccine is likely to include many elderly persons the phenomenon of immunosenesence, a reduction of immune responses in old age, must be considered. For example, poor responses to vaccines in the elderly has been noted for influenza vaccines (Bernstein *et al.*, 1999). A large phase I dose ranging trial of a toxoid A/B vaccine performed by Sanofi-Pasteur (Greenberg *et al.*, 2012) examined this concern by comparing antibody response in an 18-55 age group with an >65 group. While seroconversion rates were similar in both groups after three doses (100% seroconversion for Toxin A, 75% for Toxin B), anti toxin A titres of the elderly age group were less than a third of those observed in the younger population. In addition, the younger population fully seroconverted after two doses - the third dose

was only strictly required for the elderly population. Results of Phase II trials must be awaited in order to determine whether these differences result in any significant in *in vivo* protection.

The toxin-mediated, non-invasive nature of CDI and the aforementioned importance of humoral immunity make it an ideal target for an immunotherapeutic aimed at neutralising the toxins and preventing tissue damage. The prevention of CDI in animal models has been demonstrated with polyclonal ovine antibodies raised against TcdA and TcdB fragments (Roberts *et al.*, 2012), by humanised murine IgG1 monoclonal antibodies raised against a range of TcdA and TcdA derived antigens (Davies *et al.*, 2013) and fully human monoclonal antibodies raised against inactivated TcdA and TcdB (Babcock *et al.*, 2006). The latter of these are currently undergoing Phase III clinical trial under the trade names of Actoxumab and Bezlotoxumab (Reichert, 2013).

Chapter 2 - Introduction to The Sortase Enzyme Family

2.1 Introduction to Sortase Enzymes

2.1.1 The Sortase Enzyme Family

Sortases are transpeptidase enzymes, ubiquitous in gram positive bacteria. They play a key role in assembly of the cell surface by covalently anchoring specific proteins to the cell wall peptidoglycan and by catalysing the assembly of pili (See Figure 2.1.1). Proteins destined for processing by sortases are initially secreted via the Sec apparatus of the general secretary pathway, following which sortase enzymes (which are often co-localised with the Sec apparatus (Kline *et al.*, 2009)) recognise a pentapeptide cell wall sorting signal (CWSS) at the C-terminus of the protein. The sortase enzyme cleaves the protein at positions P4 and P5 of the CWSS (Navarre and Schneewind, 1994), resulting in an acyl enzyme state. In wall-anchoring sortases, the acyl enzyme is resolved by nucleophilic attack by an amine group within the peptidoglycan, covalently anchoring the target protein to the cell wall. The identity of the nucleophile varies with the peptidoglycan structure – in species with pentaglycine cross links, it is the N-termini of cross-link precursors (Navarre, 1998), while species with diaminopimelic acid cross-links, it is the free amino group of diaminopimelic acid (Budzik *et al.*, 2008b). In pilus-associated sortases which catalyse pilin-pilin polymerisation, the nucleophile is a lysine residue contributed by a pilin subunit on the nascent pilus (Ton-That *et al.*, 2004).

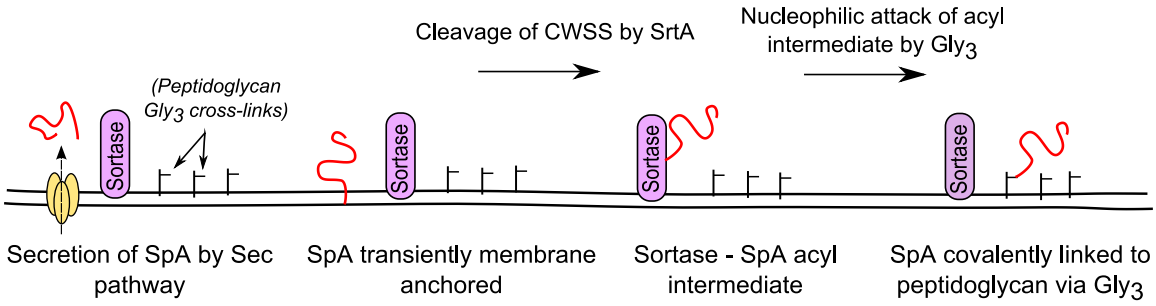


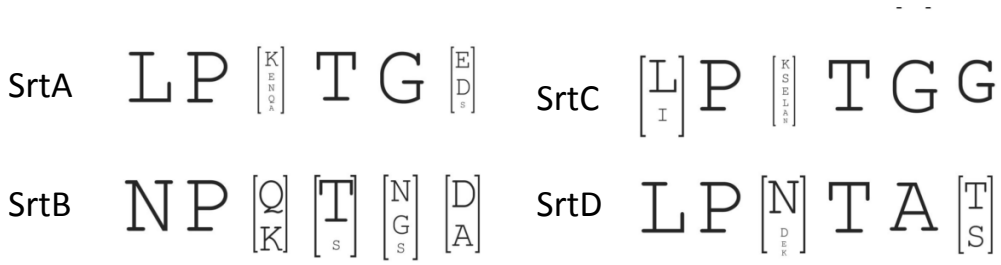
Figure 2.1.1 Schematic of Wall Anchoring of Protein A (SpA) by *S. aureus* Sortase A. Following secretion by the Sec apparatus, the target protein is cleaved by Sortase A between the threonine and glycine of its 'LPXTG' cell wall sorting signal. The resulting acyl enzyme intermediate is attacked by the polyglycine moiety of a peptidoglycan crosslink precursor, covalently anchoring SpA to the cell wall.

Sortase family proteins can be classified into five subfamilies (Dramsi *et al.*, 2005) – Sortase A, B, C, D1 and D2 (SrtA, SrtB, SrtC and SrtD1 and SrtD2).

2.1.2 The Cell Wall Sorting Signal

Sortase substrates possess three key features (Figure 2.1.2). The most essential of these is a pentapeptide CWSS motif situated 20-30 amino acids from the C terminus. Following the CWSS is a predominantly hydrophobic region of ~20 amino acids in length, while the C-terminus often contains several lysine or arginine residues resulting in a positively charged 'tail'. The purpose of the latter two features has not been investigated in depth, but there is evidence that the hydrophobic region facilitates transient membrane anchoring.

The sequence of the CWSS follows a loose consensus in most species, but contains only one strictly conserved feature, a proline residue at position two (see Figure 2.1.2). In addition to varying between species, variation of the CWSS sequence of this motif in species with multiple sortases allows proper segregation of the multiple sortase pathways.



Species/Protein	C-terminal Protein Sequence	Sortase Class
S. aureus SpA	LPETGEENPFIGTTVFGGLSLALGAALLAG**RRR**EL	(SrtA)
S. aureus IsdC	NPQTNAGTPAYIYAIPVASLALLIAITLFVR**KK**S**K**GNVEQ	(SrtB)
S. pneumoniae RrgA	YPRTGGIGMLLFYLIGCMMMGGVLLYT**RK**HP	(SrtC2)
S. pneumoniae RrgB	IPQTGGIGTIIFAVAGAVIMGIAVYAYV**K**NN**K**DEDQLA	(SrtC1)
B. anthracis BasI	LPNTASNNVAMMALSACLVGIGTLFGL**KRR**N**K**V**K**A	(SrtD)

Figure 2.1.2. The Cell Wall Sorting Signal. A) Pictographic representations of the consensus sorting signals of the four classes of sortase based on *in silico* analysis of 72 species (*excerpt from Comfort and Clubb (2004)).* Position variants are shown only if they occurred in >8% of substrates. B) Selection of C-termini from proteins processed by sortases, highlighting the CWSS (underlined) and their positively charged residues (red).

2.2 Microbiology of the Sortase Enzymes

2.2.1 – The Class A Sortase Enzymes

The Sortase A (SrtA) of *S. aureus* was the first to be identified and is considered the prototype sortase. It was discovered by investigators isolating random mutants deficient in surface display of Staphylococcal protein A (SpA) (Mazmanian *et al.*, 1999). SpA is a virulence factor which binds to the Fc region of host antibodies, and hence facilitates evasion from phagocytosis and complement mediated killing (Kim *et al.*, 2012; Palmqvist *et al.*, 2002) Even before the discovery of SrtA, the anchoring of SpA to the cell surface was well understood. The feature responsible for the anchoring had been identified as an LPxTG C-terminal peptide motif, the CWSS, followed by a 20-30 hydrophobic amino acids and a positively charged C-terminus (Schneewind *et al.*, 1992). This C-terminal signal is both essential and sufficient for wall-sorting – mutant SpA which lacks this signal is secreted into the growth medium while exogenous fusion proteins appended with the signal are anchored to the cell wall like SpA. In 1999, Mazmanian and coworkers reported a mutant deficient in SpA anchoring – the mutation responsible was mapped to a open reading frame encoding a 206 amino acid protein with a predicted N-terminal secretion signal.

There are several other *S. aureus* virulence factors which are wall anchored by SrtA, including the fibrinogen-binding clumping factors ClfA and ClfB (McDevitt *et al.*, 1994) and the fibronectin binding proteins FnbA and FnbB (Flock *et al.*, 1987). *S. aureus* mutants deficient in SrtA fail to bind IgG, fibronogen or fibrinonectin, and are severely attenuated in virulence (Mazmanian *et al.*, 2000).

In Listeria monocytogenes, 13 SrtA substrates have been identified, including three internalins, three peptidoglycan hydrolases and a 5' nucelotidase (Pucciarelli *et al.*, 2005). In *Streptococcus gordonii*, a human oral commensal, eight SrtA wall-sorted proteins have been identified (Davies *et al.*, 2009) including the adhesins CshA, CshB, SspA, SspB, a putative 5' nucelotidase and several proteins of unknown function, one of which displays homology to the 'B' domains of the *S. aureus* collagen adhesin Cna. *S. gordonii* SrtA deletion mutants are deficient in binding to salivary pellicles (Davies *et al.*, 2009). The *S. pyogenes* SrtA is responsible for cell wall anchoring of the M protein, a protease regulator, the fibronectin binding protein Protein F, and the ScpA protease which cleaves C5a of the compement cascade (Barnett and Scott, 2002)

Despite the characterisation of SrtA as a housekeeping sortase, in some species which possess SrtC-assembled pili, such as *Streptococcus agalactiae*, it is the housekeeping SrtA which anchors the first pilus subunit to the peptidoglycan (Necchi *et al.*, 2011). Indeed, the *Corynebacterium*

diptheriae SrtA catalyses assembly of an entire pilus. (Ton-That *et al.*, 2004).

2.2.2 – The Class B Sortase Enzymes

The first class B sortase to be characterised was that of *S. aureus*, following identification of an open reading frame with homology to *srtA* within the iron-responsive surface determinant (*isd*). locus. Mutagenesis of the *srtB* gene resulted in loss of wall anchoring of the adjacently encoded IsdC protein, which contains an 'NPQTN' CWSS (Mazmanian *et al.*, 2002), and a near iron transporter (NEAT) haemeoprotein binding domain (Dryla *et al.*, 2003)

Upstream of the *isd* locus promoter is a binding site for the ferrichrome uptake repressor (Fur). Consequently, growth of *S. aureus* in iron-limited medium increases expression of SrtB, as does deletion of the *fur* gene (Mazmanian *et al.*, 2002). In this species, SrtB and SrtA appear to operate cooperatively in iron acquisition, as two of the genes encoded within the *isd* locus, IsdA and IsdB, contain 'LPXTG' SrtA CWSSs rather than 'NPQTN' SrtB sorting motifs.

In addition to *S. aureus*, class B sortases have been characterised in *Bacillus anthracis* (Maresso *et al.*, 2006), *Listeria monocytogenes* (Bierne *et al.*, 2004) and *Streptococcus pyogenes*. In *L. monocytogenes*, *srtB* is under the control of a *fur* box as is the case in *S. aureus* (Newton *et al.*, 2005) and proteomic analysis of a sortase knockout strain identified two sortase B substrates, Lmo2185 and Lmo2186 possessing CWSS motifs of 'NAKTN' and 'NPKSS' respectively (Mariscotti *et al.*, 2009) These proteins are predicted to contain near-iron transporter (NEAT) domains, and *L. monocytogenes* SrtB mutants are deficient in uptake of 59Fe-labelled haemin (Xiao *et al.*, 2011). *B. anthracis* SrtB is responsible for the anchoring of a single haem binding protein IsdC (Maresso *et al.*, 2006), and deletion of either *srtB* or *isdC* results in defective utilisation of haemin.

Thus in the species so far characterised, SrtB fulfils a specialised role in iron acquisition systems, and particularly scavenging of iron from host proteins. Iron homeostasis is a critical function for any pathogen, since the concentration of free iron in human and animal tissues is virtually nil with most iron being bound to transporter proteins such as haemoglobin and transferrin.

2.2.3 The Class C Sortase Enzymes

Many Gram-positive bacteria possess pili, extracellular appendages constructed by polymerisation of multiple protein subunits, known as pilins. Pili are typically composed of major pilins, which form the main shaft, and minor or 'accessory' pilins, which often have virulence-associated effector functions. In several species, it has been demonstrated that the covalent linkage of these components is catalysed by sortases, particularly those of the class C. In sortase catalysed pilin

polymerisation, the sortase enzyme cleaves a C-terminal CWSS similar to that for SrtA, but rather than the acyl-enzyme intermediate being resolved by nucleophilic attack by a component of the peptidoglycan, the attack is from the epsilon amine of a lysine residue within a conserved motif of a second pilin at the terminus of the nascent pilus (Budzic 2008, Mandlik 2008).

Unlike SrtA and SrtB, where only one gene copy is typically present per genome, it is common for bacteria to possess multiple SrtC genes (Dramsi *et al.*, 2005), each encoding an enzyme with a specialised role in the pilin biogenesis pathway. For example, the *S. pneumoniae* pilus gene locus *rlrA* encodes three class C sortases (SrtC1, SrtC2, SrtC3), three pilins (RrgA, RrgB, RrgC) and a RofA-like transcription factor (Barocchi *et al.*, 2006). In this system, the main fibre of the pilus is composed of RgrB and polymerised by SrtC1 (El Mortaji 2011). RgrA, a collagen adhesin (Izoré *et al.*, 2010) which binds to epithelial cells *in vivo*, forms accessory fibres assembled by SrtC2 and anchored to the pilin fibre by SrtC3 (El Mortaji *et al.*, 2011).

Not all sortase-associated pili require multiple sortases. The pili of *Corynebacterium diptheriae*, for example, are composed of major pilin SpaA and the accessory pilins SpaB and SpaC, and are assembled by a single SrtC (Khare *et al.*, 2011). In *B. anthracis*, SrtC (often termed SrtD due to genome annotation conventions) catalyses the transpeptidation reaction between the LPxTG sorting signal of the major pilin BcpA and the gamma amino group of lysine of an YPKN motif on an adjacent BcpA unit (Budzik *et al.*, 2008a). The *B. anthracis* SrtA is also capable of cleaving the LPxTG signal, and is responsible for anchoring of the first subunit of the pilus to the peptidoglycan (Budzik *et al.*, 2007). The minor pilin BcpB is also cleaved by SrtD but because it lacks a YPKN motif, addition of this subunit to the pilus effectively terminates further pilus growth (Budzik *et al.*, 2008a).

2.2.4 – The Class D Sortase Enzymes

Class D sortases are the least investigated of all classes, with characterisation limited to the species *B. anthracis*. The CWSS recognised by *B. anthracis* SrtD (termed SrtC in the literature due to genome annotation conventions) is very similar to that of its SrtA, specifically an 'LPNTA' sequence as opposed to 'LPxTG' recognised by *B. anthracis* SrtA. The two proteins processed by *B. anthracis* SrtD are BasH and BasI (Marraffini and Schneewind, 2006), both of which are targeted to the diaminopimelic acid crosslinks of the peptidoglycan of cells undergoing sporulation (Marraffini and Schneewind, 2007). Their precise function is as yet unknown, but deletion of SrtD in *B. anthracis* results in a sporulation-deficient yet fully virulent strain (Marraffini and Schneewind, 2006).

2.3 Biochemistry of the Sortase Transpeptidation Reaction

2.3.1 Overview and Reaction Kinetics

The sortase reaction is of the 'ping-pong bi-bi' type (Huang *et al.*, 2003), where 'bi-bi' describes that there are two substrates and two products while 'ping-pong' mechanisms are so named because the enzyme exists in one of two states, either free or in a substrate-modified state. Only the substrate-modified state is competent for binding of the second substrate, hence the overall reaction proceeds in a highly ordered manner. In the case of the sortase enzymes, this substrate-modified state is the acyl-enzyme complex formed by a nucleophilic attack of the enyzme upon the threonyl-glycyl bond of the CWSS.

Kinetic data for SrtA collected at a fixed concentration of substrate 2 (S2) and varying concentrations of substrate 1 (S1) do not fit the expected parameters for a ping pong bi bi mechanism – rather, they fit the parameters for a ping pong reaction incorporating a hydrolytic shunt (Frankel *et al.*, 2005). In such a mechanism, water competes with substrate 2 in the second part of the reaction, so that at low concentrations of substrate 2, the enzyme is returned to its free state via hydrolysis reaction rather than transpeptidation (See Fig. 2.3.1). Accordingly, in the absence of a suitable nucleophile to complete the transpeptidation reaction, hydrolysis of the CWSS is observed (Huang *et al.*, 2003).

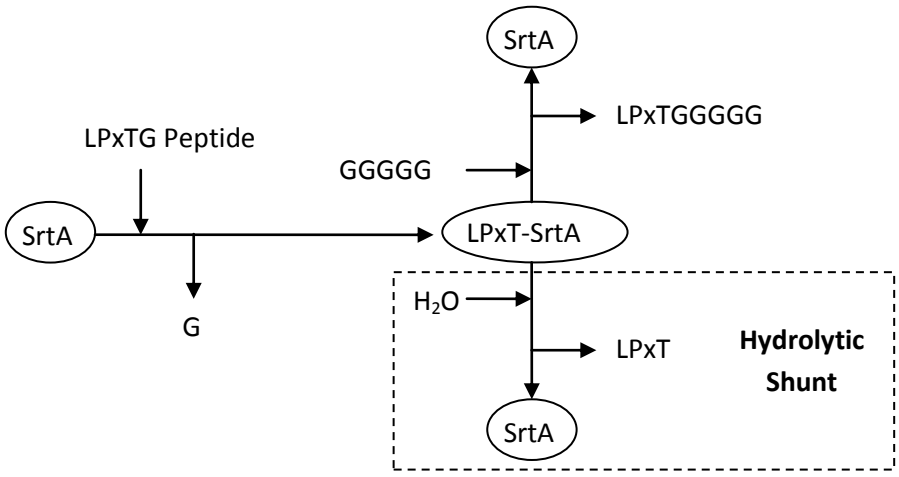


Figure 2.3.1 – Ping-Pong Transpeptidation by SrtA Incorporating a Hydrolytic Shunt. Cleavage of the LPxTG motif to generate the LPET-SrtA acyl enzyme releases a glycine residue. From this acyl state, the return of the enzyme to its free state can proceed *via* either transpeptidation with pentaglycine (GGGGG), producing an 'LPETGGGGG' peptide, or hydrolysis, releasing an 'LPET' peptide.

2.3.2 Mechanism of the Attack on the Threonyl-Glycyl Bond of the CWSS

The rate limiting step in sortase transpeptidation is formation of the acyl-enzyme (Huang *et al.*, 2003). The mechanism of this reaction, so far only investigated in *S. aureus* SrtA, has been revised several times recently in light of new evidence from a variety of structural and biochemical experiments.

The first structure of *S. aureus* sortase A (Ilangovan *et al.*, 2001) revealed that the conserved and essential residues Histidine 120 (His120) and Cysteine 184 (Cys184) (Ton-That *et al.*, 2002) were in close proximity (for detail, See Section 2.5.2). It was therefore initially proposed that SrtA operated in a mechanism analogous to that of the cysteine protease papain, via a thiolate-imidazolium ion pair. There is, however, strong biochemical evidence against a thiolate imidazolium ion pair in SrtA. In papain, the ion pairing results in a pKa of the catalytic histidine residue as low as 3.3-4.0 (Lewis *et al.*, 1981; Pinitglang *et al.*, 1997), whereas in SrtA the pKa of His120 has been determined as 7.0 (Connolly *et al.*, 2003), and hence does not allow for the low level of protonation as is required by a thiolate imidazolium pair model. Additionally, replacement of Cys184 with Ala has very little effect on the pKa of His120 in SrtA, whereas neutralisation of the papain catalytic cysteine by methylthiolation leads to a dramatic increase in the pKa of the paired His from 4 to 8.5 (Lewis *et al.*, 1981). It therefore appears His120 and Cys184 of SrtA do not participate in an ion pair.

An alternative mechanism for activation of Cys184 was therefore proposed, a general base activation mechanism whereby Cys184 exists in thiol form in the resting state, and is activated by abstraction of a proton by a base, proposed to be either Arg233 or His120. By determining k_{cat} and k_{cat}/K_M values over a range of pH values, Frankel *et al* (2005) determined that enzyme activity was dependent upon two ionisable groups, with pKa values of 6.3 and 9.4, assigned to His120 and Cys184 respectively. Significantly, no effect was observed at or near that expected for an arginine residue as would be expected if Arg197 was acting as a general base. Additionally, mutation of Arg197 to either alanine or lysine does not change the pH-k_{cat} curve (Frankel *et al.*, 2007a), confirming that Arg197 does not act as general base to activate Cys184. A variation on the general base mechanism is that of a Cys-His-Asp catalytic triad where protons abstracted by the histidine residue are shuttled by Asp. This proposition was based on structural evidence of a catalytic triad similar to that seen in serine and cysteine proteases (Zhang *et al.*, 2004). However, mutation of Asp185 and Asp186 has little effect on enzyme activity (Frankel *et al.*, 2007a) .

Based on the above evidence, Frankel *et al* (2007) proposed a reverse protonation mechanism which does not implicate any special activation of Cys (Frankel *et al.*, 2005). At pH7, a small proportion (estimated at less than 0.06%) of SrtA molecules are 'reverse protonated', ie. Cys is in the thiolate form while histidine is in the imidazolium form – a configuration which occurs, albeit rarely, due to overlaps in the protonation curves of histidine and cysteine at pH 7. The reverse protonation mechanism is compatible with pKa measurements, mutagenesis experiments, solvent isotope effects, and explains several unusual features such as the unusually low k_{cat} of SrtA (Kruger *et al.*, 2004b). The proposed role of His120 in this model is to donate a proton to the leaving amine group.

2.3.3 Role of the Conserved Arginine Residue in Oxyanion Stabilisation

The reverse protonation model for sortase catalysis does not directly implicate the conserved arginine residue, yet it is essential for optimal enzyme activity, with mutation of Arg197 to alanine resulting in a 10^3 decrease in k_{cat} (Frankel *et al.*, 2007b). It was therefore proposed that the role of Arg197 is to stabilise the oxyanion of the tetrahedral intermediate formed after the attack upon the threonyl glycyl bond. The calculated energetic contribution of Arg197 to oxyanion stabilisation is similar to that seen in other enzymes which employ oxyanion stabilisation, eg. carboxypeptidase A.

Structural evidence from SrtB is compatible with Arg197 fulfilling an additional role in activating the nucleophile which attacks the acyl enzyme (see Section 2.5.4).

2.3.4 Substrate Specificity

The CWSS specificity of *S. aureus* SrtA has been comprehensively investigated by the use of a peptide mutant library (Kruger *et al.*, 2004d). In an experiment examining initial reaction rates, only the canonical amino acids were tolerated at P1, P2, P4 and P5 (LPTG respectively), while P3 was tolerant of any amino acid tested (ADEFGILMNPQSTYV). The relative rates of P3 variants do not appear to be correlated to side chain side size or polarity, suggesting that this residue undergoes minimal interaction with the enzyme. In an end-point experiment over a longer period of time, it was observed that SrtA is partially tolerant of methionine at P1, of alanine at P2, and of alanine, leucine, serine and valine at P4. However, glycine remained the only amino acid tolerated at P5.

Because the acylation step of the reaction is rate limiting only in the presence of an appropriate nucleophile (Huang *et al.*, 2003) specificity for the nucleophilic transpeptidation substrate can be assessed by comparison with hydrolysis rates. Of the pentaglycine crosslink precursor in the *S. aureus* peptidoglycan, diglycine is the minimum required unit for transpeptidation (Huang *et al.*,

2003). SrtA is highly selective for glycine at P1, while positions P2 and P3 are more tolerant, with GlyVal and GlyAla both tolerated albeit with 5- and 10-fold increases in K_M. Addition of GlyGlyHis and GlyGlyLeu resulted in reaction parameters very similar to those where tetraglycine is added. Addition of monomeric glycine does not result in transpeptidation, but glycinamide does, with reaction parameters similar to those of triglycine.

2.3.5 Sortase Inhibitors

Since the discovery of SrtA as an essential virulence factor in *S. aureus* (Mazmanian *et al.*, 2000), the development of sortase inhibitors as an anti-infective treatment has received great interest, and several candidate inhibitors have been investigated. The first reported inhibitor of SrtA was 2-(trimethylammonium) ethyl methanethiosulfonate (MTSET), a thiol reactive compound which reacts irreversibly with cysteine thiol of the sortase active site (Zong *et al.*, 2004b). However, the short half-life and broad reactivity of this compound renders it useful only as an experimental *in vitro* inhibitor (Stauffer and Karlin, 1994).

The CWSS substrate is relatively easy to mimic due to its peptide nature, and thus several substrate mimetic compounds have been developed. For example, two inhibitors of *S. aureus* SrtA with sub-micromolar inhibition constants were developed by replacement of the C-terminal glycine with either a diazoketone or chloromethyl group (Scott *et al.*, 2002), both of which have been found to alkylate active site cysteine residues in other enzymes. Of the two compounds tested, the chloromethyl-substituted peptide modified the enzyme at twice the rate of the diazoketone peptide although still slow, at a rate of $1.1 \times 10^{-2} min^{-1}$. Replacement of the scissile Thr-Gly bond with a nonhydrolysible phosphinic bond also inhibited *S. aureus* SrtA but with a poor Ki of 11.4mM (Kruger *et al.*, 2004a).

In addition to substrate mimetics, several small molecule inhibitors have been identified. High throughput screening of sortase inhibition against a 135,635 small molecule library revealed 207 specific inhibitors of *B. anthracis* SrtA (Maresso *et al.*, 2007). Of these, several aryl (β-amino) ethyl ketones (AAEKs) were investigated further by X-ray crystallography of the inhibited enzyme, by which it was found that the inhibition was due to covalent adduction at the active site cysteine residue.

To date, there is only one published report of *in vivo* testing of a sortase inhibitor, in which it was demonstrated that (Z)-3-(2,5-dimethoxyphenyl)-2-(4-methoxyphenyl) acrylonitrile (DMMA) reduced mortality from *S. aureus* in mouse models of kidney and joint infection (Oh *et al.*, 2010).

2.4 The Structure of the Sortase Family Enzymes

2.4.1 Overview of Sortase Structures

The first published sortase structure was that of *S. aureus* SrtA (Figure 2.4.1), determined by NMR spectroscopy (PDB 1IJA, Ilangovan *et al.* 2001). This revealed a novel structure composed of an eight stranded beta-barrel, flanked by one short alpha helix and one 3_{10} helix. The unique sortase fold is formed by the interaction of two structural motifs, each composed of four strands. One strand from each motif (β4 and β7) interact along much of their lengths, while the remaining strands (β1,β2,β3 and β8,β6,β5) co-operatively form the remainder of the barrel which is completed by the interactions of β1 and β5.

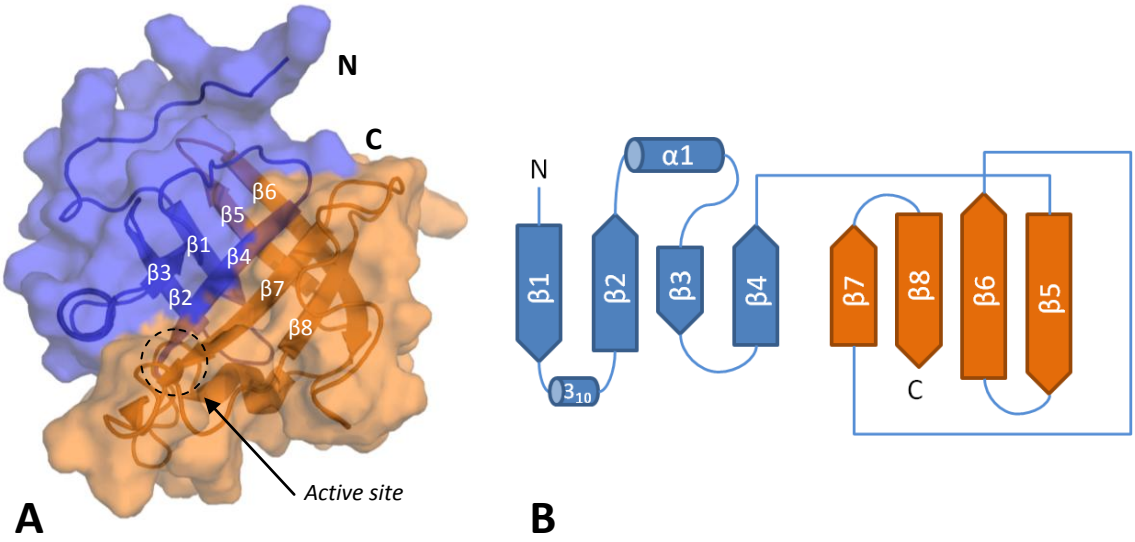


Figure 2.4.1. Structure of *S. aureus* SrtA. (A) Cartoon rendering of the NMR structure of *S. aureus* SrtA (PDB 1IJA, Ilangovan *et al*, 2001) (B) Schematic representing the secondary structure and topology of 1IJA.

Subsequent X-ray structures of *S. aureus* SrtA are in broad agreement with the NMR structure (eg. PDB 1T2P, Zong *et al.* 2004), and crystal structures of several other SrtA enzymes have been determined, including those of *Bacillus anthracis* (PDB 2KW8, Weiner *et al.* 2010), *Streptococcus pyogenes* (PDB 3FN7, Race *et al.* 2009) and *Streptococcus agalactiae* (PDB 3RCC, Khare et al. 2011).

The first structure of a class B sortase, again that of *S. aureus*, was determined by X-ray crystallography (See Fig 2.4.2). The core beta barrel structures of *S. aureus* SrtA and SrtB are very similar despite only sharing 40% sequence similarity. For the alpha carbon atoms of the β-strand core regions, Zong *et al.* (2004) report a root-mean square deviation (RMSD) of only 1.55Å.

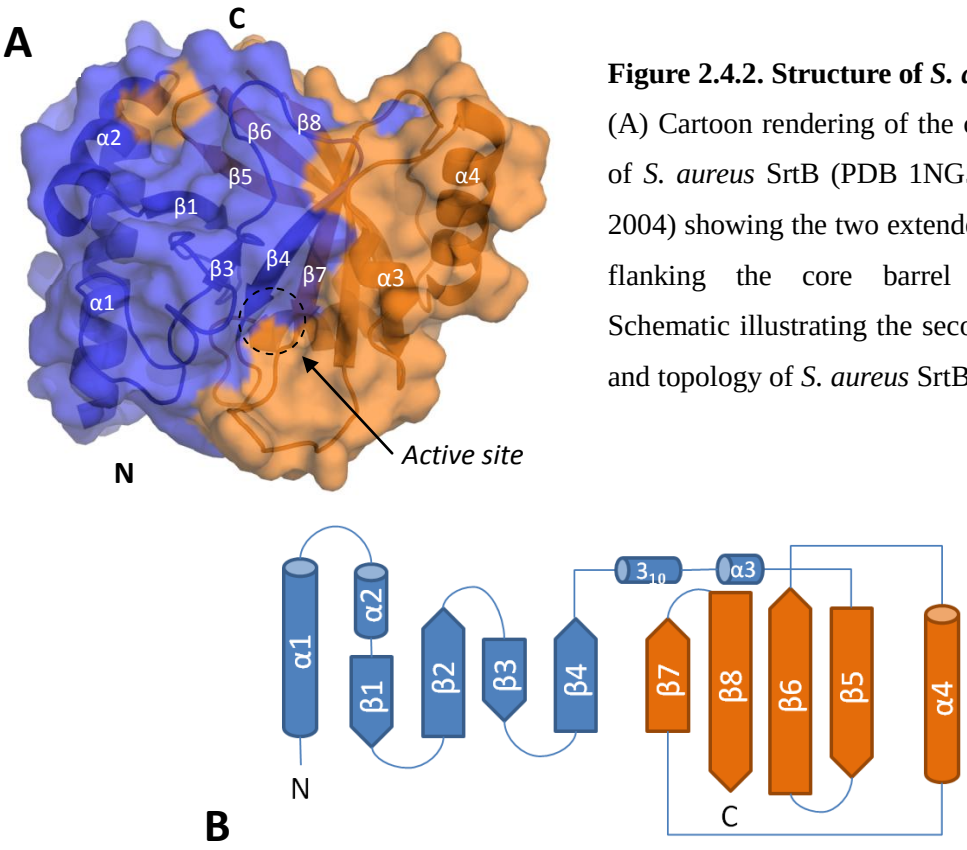


Figure 2.4.2. Structure of *S. aureus* SrtB. (A) Cartoon rendering of the crystal structure of *S. aureus* SrtB (PDB 1NG5, Zhang *et al.*, 2004) showing the two extended alpha helices flanking the core barrel structure (B) Schematic illustrating the secondary structure and topology of *S. aureus* SrtB.

Despite the similarities between the core structures of *S. aureus* SrtA and SrtB, there are significant structural differences outside of the barrel region. Five helical regions are apparent in SrtB, most notably long alpha helices at the N-terminus and in the β6-β7 loop. Two further short alpha helices are present in the β4-β5 loop, and a short region following the N-terminal helix, separated from it by a turn motif. The presence of additional helices in SrtB relative to SrtA has the effect of placing the N-terminal membrane anchor proximally to the active site as opposed to SrtA where it is distal to the active site. This results in the SrtA active site being oriented away from the cell surface while the SrtB active site is oriented towards the surface. The significance of this difference has not been investigated, but may result in differences with respect to access to the peptidoglycan.

Another significant difference between SrtA and SrtB is the configuration of the β7-β8 loop, which is much longer in SrtB than in SrtA. However, the relative positioning of the catalytic cysteine and conserved arginine residues at either end of this loop is well conserved. The structures of SrtB enzymes from two other species have been published, *S. pyogenes* (PDB 3PSQ, Kang 2011) and *B. anthracis* (PDB 1RZ2, Wu 2004).

Several class C Sortase structures have been determined by X-ray crystallography. Most notably, the structures of all three *S. pneumoniae* SrtCs have been solved (PDB 2W1J, 3G66, 2W1K, Manzano *et al.*, 2008, Neiers *et al.*, 2009) providing a complete structural overview of pilus biogenesis in this species. In common with SrtA and SrtB, the core SrtC structure is that of an 8 stranded beta barrel, and as in the case of SrtB the core is flanked by several helices (See Figure 2.4.3).

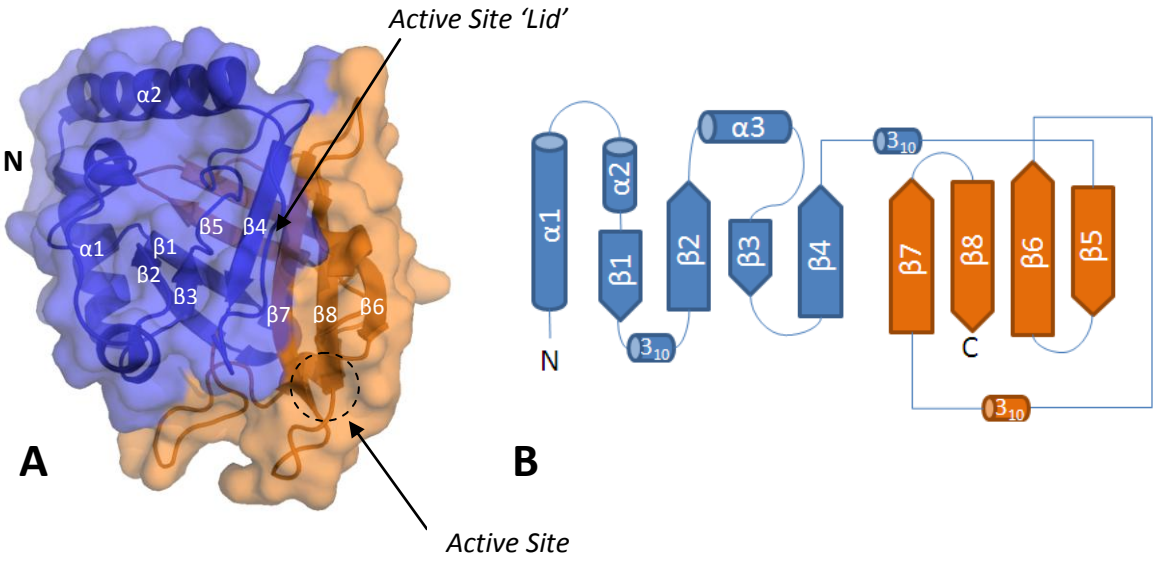


Figure 2.4.3. Structure of *S. pneumoniae* SrtC1. (A) Cartoon rendering of *S. pneumoniae* SrtC1 (PDB 3G66, Neiers *et al.*, 2009) showing the loop which forms a lid occluding the active site (B) Schematic illustrating the secondary structure and topology of 3G66.

The key distinguishing structural feature of class C sortases is a loop which forms a 'lid' occluding the active site. In crystal structures of the three *S. pneumoniae* SrtCs, the lid region exhibited high B-factors compared to the remainder of the structures, suggesting a high level of flexibility, and similarly, in the case of SrtC2, the lid was observed in different conformations in two crystals grown in different conditions. In the *Streptococcus agalactiae* SrtC1, no electron density was observed in the loop region, again suggesting a high level of mobility (Khare *et al.*, 2011).

In *S.pnuemoniae*, the lid is anchored by a conserved salt bridge between Asparagine 73 within the lid and Arginine 215 within the β7-β8 loop (Manzano *et al.*, 2008). Similar pairings are observed in other species eg *Actinomyces oris* and *S. agalactiae* (Khare *et al.*, 2011; Persson, 2011). This is particularly interesting given that this arginine residue is conserved in all sortase classes, and it has been suggested that the role of this interaction is not only to lock the lid in place, but also to regulate positioning of Arg215 for catalysis (Persson, 2011) . The role of the lid may be to regulate

substrate access, or it may serve to protect against non-specific oxidation of the active site sulphydryl. Despite its uniqueness to pilin-associated sortases, the role of the lid is clearly not specific to catalysis of pilin-pilin reactions, as *S. pneumoniae* SrtC3 possesses the lid despite only being able to catalyse pilin-peptidoglycan transpeptidation, and SrtC1 and SrtC2, both possessing the lid, can catalyse either pilin-peptidoglycan or pilin-pilin transpeptidation.

Class D sortases are the least well characterised of the four classes and only one structure has been published, that of the *B. anthracis* SrtD determined by NMR spectroscopy (often named 'SrtC' due to naming conventions in the *B. anthracis* genome). *B. anthracis* SrtD is very closely related to SrtA, and can be superposed on the *S. aureus*, *B. anthracis* and *S. pyogenes* SrtA structures with carbon RMSD values of 2.4Å, 1.9Å and 2.0Å respectively (Robson *et al.*, 2012) hence has not been illustrated here.

2.5.2 Structure of the Sortase Active Site

The active site of SrtA resides within a concave depression formed of β4 and β7, which provide the floor of the depression, and β2/β3, β3/β4 and β6/β7 which form the sides of the cleft (See Fig. 2.4.4). Located at one end of this concave surface, within β7, is the conserved 'TLxTC' motif containing the catalytically essential cysteine residue. Flanking this residue on adjacent loops are the conserved histidine and arginine residues, the putative functions of which were discussed in Section 2.3.2.

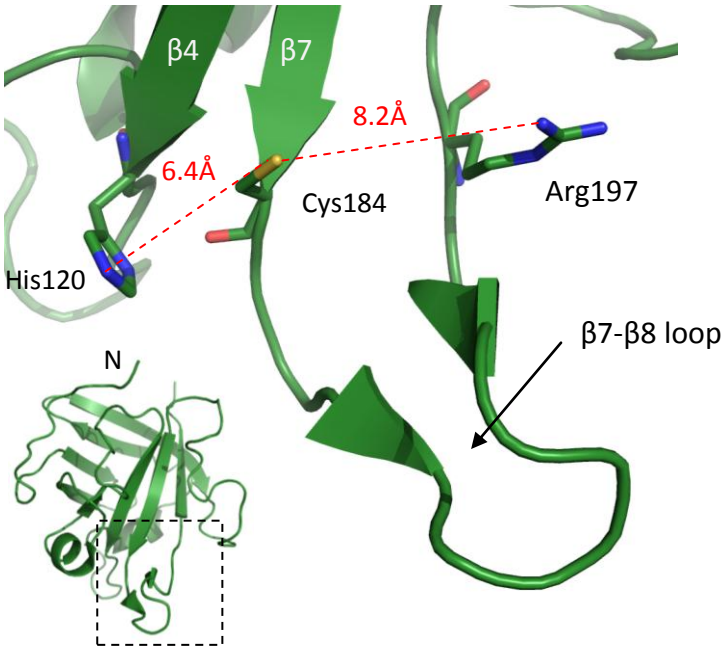


Figure 2.4.4 - The Active Site of *S. aureus* SrtA (PDB 1IJA, Ilangovan *et al*, 2001). The conserved and essential Cys184 residue at the C-terminus of β7 is flanked by conserved His120 and Arg197 residues on adjacent loops at distances of 6.4Å and 8.2Å respectively.

2.4.3 Structural Evidence for Recognition of the CWSS

S. aureus SrtA binds the LPxTG CWSS via strands β4, β7 and the loops which flank them, particularly β6-β7 (Suree *et al.*, 2009). Underlining the role of this loop in substrate recognition, targeted mutations within the β6-β7 loop can confer substrate promiscuity to SrtA (Piotukh *et al.*, 2011), while a loop swap mutant consisting of SrtA with the β6-β7 loop of SrtB confers SrtB-like substrate recognition (Bentley *et al.*, 2007). Comparison of NMR spectra of *apo* SrtA and SrtA-LPAT* (SrtA modified with an irreversible inhibitor mimicking an LPETG sorting signal) reveals chemical shift perturbations in β4 (A118), β7 (I182), β3-β4 (A104, E105), β6-β7 (L169, E171) and β7-β8 (V193, W194, R197) (Liew *et al.*, 2004). Many of these interactions are depicted in Figure 2.5.5.

Further analysis of SrtA-LPAT* indicates that binding of the sorting motif to the β6-β7 loop may be via an induced fit mechanism. In the *apo* structure, loop β6-β7 is highly mobile and largely free of secondary structure (see Figure 2.4.6). Upon substrate binding, residues within β6-β7 (V166, V168, L169) rearrange to form a hydrophobic 3_{10} helix which interacts extensively with the leucine and proline residue of the CWSS, providing a stronger hydrophobic interaction (Suree *et al.*, 2009). Proline, the second residue of the CWSS, is observed interacting with I182, A118, L169, A92 and A104.

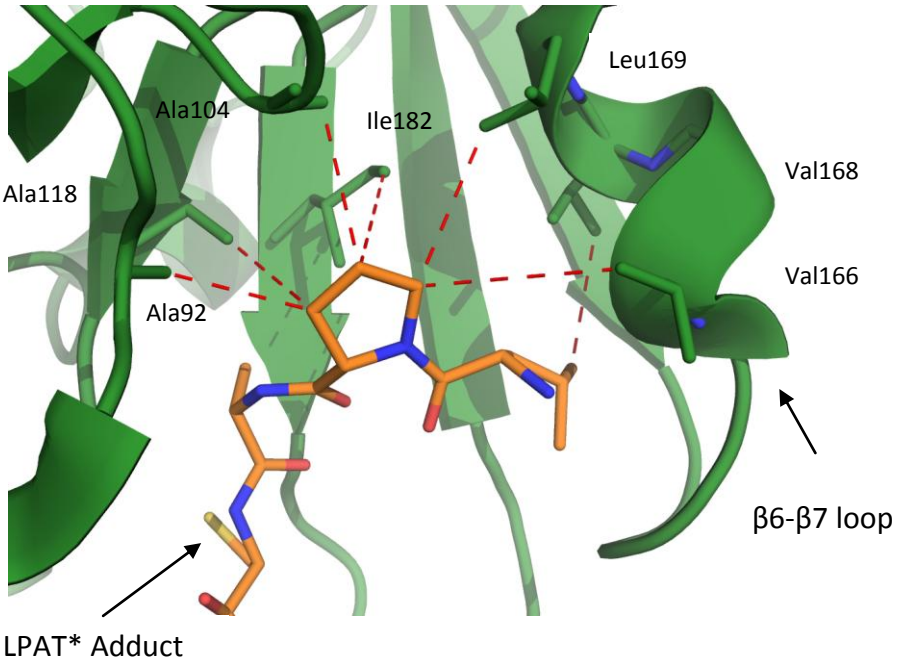


Figure 2.4.5 – Structure of *S. aureus* SrtA with a Substrate Mimetic Adduct (PDB 2KID, Suree *et al.*, 2009). The proline and leucine residues of the LPxTG CWSS are placed to form extensive hydrophobic interactions (represented here by red dashed lines) with residues surrounding the substrate binding cleft.

Underlining the importance of hydrophobic interactions in the active site, mutation of Thr180, Leu181 or Ile182 to alanine result in 14, 7.6 and 28-fold reduction in enzyme activity respectively, and mutation of Ile182 to serine, a more polar amino acid, results in a 74-fold reduction in activity (Frankel *et al.*, 2007b).

The solution structure of *B. anthracis* SrtA, despite being free of substrate, displays a helix within the β6-β7 loop, similar to that observed in *S. aureus* SrtA-LPAT* (see Fig. 2.4.6). The *B. anthracis* SrtA β6/β7 loop may not, therefore, require a transition to order upon substrate binding, as is the case in *S. aureus* SrtA. A similar pre-formed helix is observed in the *S. pyogenes* SrtA, although the high B factors observed within the helix suggest that the loop is indeed mobile.

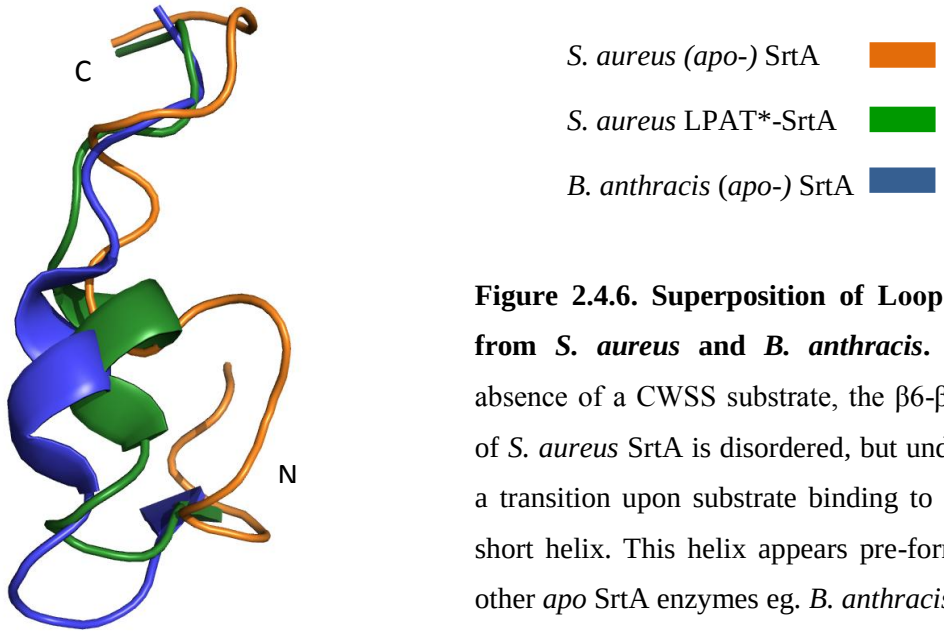


Figure 2.4.6. Superposition of Loop β6-β7 from *S. aureus* and *B. anthracis*. In the absence of a CWSS substrate, the β6-β7 loop of *S. aureus* SrtA is disordered, but undergoes a transition upon substrate binding to form a short helix. This helix appears pre-formed in other *apo* SrtA enzymes eg. *B. anthracis* SrtA.

In addition to the NMR structure of SrtA-LPAT*, a crystal structure of SrtA Cys184 to Alanine mutant complexed with an LPETG peptide is available (PDB 1T2W, Zong *et al.*, 2004). However, this is quite different to the structure of SrtA-LPAT* presented by Suree *et al* and displays some features which suggest that the peptide is bound non-specifically. For example, the leucine side chain of the LPETG peptide is oriented away from the enzyme and towards the solvent, contradicting biochemical data which indicate that there are specific enzyme-substrate interactions at this position (Kruger *et al.*, 2004d). Similarly, the $SrtA_{C184}$-LPETG structure shows a hydrogen bond between Q172 and the backbone of the LPETG peptide, whereas biochemical data from Q172A mutants (Bentley *et al.*, 2008) indicate that this residue is dispensable for catalysis so is unlikely to participate in such a strong interaction with the substrate.

At present, there are no published structures of a class B sortase complexed with a sorting motif peptide. As discussed, the β6/β7 loop of *S. aureus* SrtA participates in significant hydrophobic interactions which anchor the 'LP' portion of the LPETG sorting signal. The analogous region in the β6/β7 loop of SrtB is more polar, reflecting the more polar nature of asparagine at P1 of the 'NPQTN' *S. aureus* SrtB sorting motif. Residues which form key hydrophobic interactions with proline in SrtA-LPAT* structure do not appear to be conserved in *S. aureus* SrtB either, with isoleucine 182 replaced with a serine residue and alanine 118 with a tyrosine residue. The active site of *S. aureus* SrtB is therefore less hydrophobic and enzyme-substrate interactions are likely to bemore polar in nature.

Neither the *B. anthracis* nor the *S. aureus* SrtB structures display a pre-formed helix within β6/β7, suggesting that they either undergo similar changes upon substrate binding as *S. aureus* SrtA or that such rearrangements are not necessary for binding of the SrtB CWSS.

2.5.4 Interaction of Sortase Enzymes with the Peptidoglycan Nucleophile

Although there are currently no published crystal structures of SrtA complexed with the polyglycine moiety which acts as a nucleophile in the transpeptidation reaction, this interaction has been partially investigated by NMR (Suree *et al.*, 2009). In complexes of SrtA-LPAT* complexed with Gly_3, perturbations were observed in residues within H1, β7/β8 and β4/H2. These regions form a surface which is exposed by a ~13Å displacement of the β7-β8 loop on the formation of an acyl intermediate. Furthermore, it was found that SrtA does not bind Gly_3 at all in its *apo* state. For *S. aureus* SrtA, the displacement of the β7/β8 loop after formation of the acyl intermediate therefore appears to be the structural basis for the ordered binding of substrates observed in bisubstrate kinetic characterisations.

A structure of *S. aureus* SrtB complexed with Gly_3 has been obtained by X-ray crystallography (1QXA, Zong et al 2004) and in agreement with NMR data for SrtA, the Gly_3 binding region is located within the β7-β8 loop (see Figure 2.5.7). Interestingly, in this structure Gly_3 is bound to SrtB in the absence of a sorting motif substrate, contradictory to the findings of Suree *et al* (2009) that *apo* SrtA is not competent for binding of the second substrate.

In order for the Gly_3 peptide to participate in a nucleophilic attack upon the thioacyl intermediate form of enzyme, it must be activated by removal of a proton from its N-terminal amino group. The highly conserved residue Arg223 is observed within 4.5Å of the N terminal amino group of Gly_3, suggesting that this is the most likely residue to fulfil the role of base activation. The suggestion that the conserved histidine is responsible for this role is not ruled out, but appears less likely as it is further away than the arginine residue at a distance of 6.4Å.

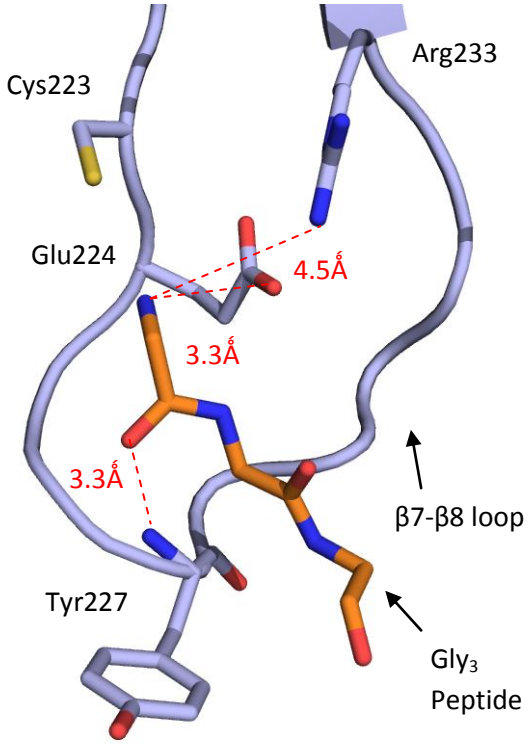


Figure 2.4.7. Interaction of SrtB with Gly3 Peptide. In the structure of a complex between *S. aureus* SrtB and a triglycine molecule (PDB 1QXA, Zong *et al.*, 2004), Gly_3 is observed within the β7-β8 loop. Distances consistent with weak hydrogen bonding are observed between: 1) The N-terminal amino group and the side chain of Glu224 and the 2) The carbonyl of the P1 Gly and the backbone amino group of Tyr227 located at the bottom of the loop. The distance between Arg233 and the N-terminus of the Gly3 peptide is 4.5Å, consistent with its putative role in activation of Gly3 prior to nucleophilic attack upon the acyl enzyme.

2.5.5 Binding of Metal Ions by *S. aureus* SrtA

Following the observation that *S. aureus* SrtA activity is stimulated eightfold by calcium ions, SrtA NMR spectra were recorded in the presence and absence of Ca^{2+}. A putative Ca^{2+} binding pocket was identified in loop β3-β4 comprising the residues Glu105, Glu108 and Asp112 (Ilangovan *et al.*, 2001). Ca^{2+} induced chemical shift changes were also observed in the adjacent β6-β7 loop, and further experiments confirm that E171 within β6/β7 interacts with Ca^{2+} bound within the β3/β4 loop, confirmed by biochemical analysis in which an Glu171A mutation results in reduced Ca^{2+} stimulation (Naik *et al.*, 2006). It is therefore proposed that Ca^{2+} stimulates the *S. aureus* SrtA by stabilisation of the β6-β7 loop, a region which is a key for binding for the sorting motif, and which exhibits a high level of mobility in Ca^{2+}-free structures (Zong *et al.*, 2004a)

However, stimulation by calcium ions is not universal in SrtA enzymes and *S. aureus* SrtA may indeed be an exceptional case. The *Streptococcus pyogenes* SrtA, is mildly inhibited by Ca^{2+}, and correspondingly, the crystal structure of this enzyme does not display a calcium binding pocket within β3/β7, although an aspartic acid residue (D196) is present in the equivalent position to Glu171 within loop β6/β7 (Race *et al.*, 2009). Similarly, the *B. anthracis* SrtA has a structurally conserved glutamic acid residue within β6/β7 but no binding pocket within β3/β4. The activity of *S. agalactiae* SrtA is also independent of Ca^{2+} (Necchi *et al.*, 2011).

2.6 Introduction to the *C. difficile* Sortase

2.6.1 Current Knowledge of the *C. difficile* Sortase

Annotation of the *C. difficile* 630 genome revealed two genes (CD2718 and CD3146) encoding proteins with homology to known sortases. The latter of these is a pseudogene, containing an in-frame stop codon ; the stop codon in CD3146 occurs upstream of the nuceleotide sequence encoding the TLxTC active site motif, and the conserved arginine is also absent, meaning that any expressed fragment is very unlikely to be active.

Based upon analysis of the *C. difficile* genome using consensus searches designed to identify wall-sorted proteins, a *C. difficile* cell wall-sorting signal of SPxTG or PPxTG was initially suggested (Pallen *et al.*, 2001). A similar methodology employed within a larger scale search of bacterial genomes identified seven putative substrates in *C. difficile*, six of which contained SPxTG motif and one with a PPxTG motif (Comfort and Clubb, 2004, supplementary data - available in the form of a freely accessible database at *http://nihserver.mbi.ucla.edu/Sortase/)*

At the time of writing, there are no literature reports of any *in vitro* investigation of the *C. difficile* sortase. However, a patent application has been made in which it is claimed that the sortase is an essential gene, a claim made upon the basis that the inventors could not isolate sortase mutants and therefore assumed it fulfils an essential function (Wren, 2011). Additionally, in a publication characterising a novel *C. difficile* collagen binding protein named CbpA, the authors observed an NVQTN motif at the C-terminus and suggested that this protein may be sortase anchored (Tulli *et al.*, 2013), in spite of the *in silico* identification of SPKTG as the *C. difficile* sorting signal.

2.6.2 Aims of the work

The aims of the work are as follows:

1. To elucidate the *in vivo* role of the *C. difficile* sortase by construction and characterisation of a sortase gene knockout (Chapter 5)
2. To provide structural information on the *C. difficile* sortase and its substrates by X-ray crystallography (Chapter 6 and 7)
3. To determine the specificity and activity of the SrtB enzyme by biochemical assays and mass spectroscopy (Chapter 8), particularly with regard to the conflicting claims of Tulli *et al* (2013) and Comfort and Clubb (2004*)* regarding the identity of the CWSS.

The general and preparative methods underpinning this work are detailed in Chapter 3 (General Methods) and Chapter 4 (Purification of Recombinant Proteins).

Chapter 3 - General Methods and Protein Purification

3.1 General Microbiological Methods

3.1.1 Chemical Reagents

All chemicals were purchased from Sigma Aldrich unless otherwise stated.

3.1.2 Bacterial Strains and Plasmids

E. coli strains BL21 and 10β were purchased from New England Biolabs in a transformation-competent state. Plasmid pEXP1 (Invitrogen) is an expression vector featuring a T7 promoter and encoding an N-terminal enterokinase-cleavable hexahistidine tag. pJ414 (DNA 2.0) is an expression vector featuring a pUC origin of replication, a multiple cloning site (MCS) downstream of a T7 promoter and an ampicillin resistance gene. pET28a (Novagen/Merck) is an expression vector featuring a pBR322 origin of replication, an MCS downstream of a T7 promoter and a kanamycin resistance gene. Its MCS is constructed such that cloned genes can be fused with sequences encoding either a thrombin cleavable N-terminal hexahistidine tag, or a non cleavable C-terminal tag.

3.1.3 Growth of *Escherichia coli*

E. coli strains BL21 and 10β were routinely grown in Lysogeny Broth (LB, 10g/L tryptone, 5g/L yeast extract, 5g/L NaCl) or on LB Agar (LB + 1% agar). For expression of recombinant proteins, *E. coli* BL21 was grown in Terrific Broth (TB, 24g/L tryptone, 42g/L yeast extract, 17mM KH_2PO_4 and 72mM K_2HPO_4)

3.1.4 Transformation of *E. coli*

Transformation-competent cells were thawed on ice, mixed with 1ng of plasmid DNA and incubated on ice for 30mins. Following heat shock for 10s at 42°C, 950µl of SOC broth was added and incubation continued for 1hr at 37°C with orbital agitation at 250rpm. Volumes of 10µl, 50µl and 200µl were spread onto LB agar plates supplemented with appropriate antibiotic – 100µg/ml ampicillin for pEXP1 and pJ414 or 50µg/ml kanamycin for pET28a.

For transformation of *E. coli* CA343, electroporation was employed. Electrocompetent cells were mixed with 1ng of plasmid DNA within a pre-chilled electroporation cuvette and pulsed with 2.5kV for 4 msec using a Bio-Rad Micropulser. After addition of 250µl of SOC broth, incubation and plating was performed as above.

3.1.5 Growth of *Clostridium difficile*

C. difficile was grown within a Don Whitley MGW-1000 anaerobic chamber maintaining an atmosphere of 10% Hydrogen, 10% Carbon Dioxide, 80% Nitrogen, a temperature of 37°C and a relative humidity of 70%. Liquid cultures were grown in Supplemented Brain Heart Infusion broth (sBHI, 36g/L Brain Heart Infusion (Oxoid), 5g Yeast Extract (Oxoid), 0.5g/l L-cystiene hydrochloride). Plate cultures were grown on Facultative Anaerobe Agar (Oxoid).

Archive and working stocks of *C. difficile* were generated by inoculating 80ml of sBHI with a single colony, growing to an optical density of 1.0, addition of dimethyl sulphoxide (DMSO) to a concentration of 10% and freezing of 1ml aliquots under liquid nitrogen.

3.2 General Molecular Biology Methods

3.2.1 Agarose Electrophoresis of DNA

Plasmids, PCR products and restriction fragment solutions were mixed 6:1 with loading buffer (40% glycerol, 0.25% Orange G) and added to the wells of a 40mm x 60mm x 5mm 1% agarose gel made with tris-acetate EDTA buffer (TAE, 40mM Tris, 20mM acetic acid, and 1mM Ethylenediaminetetraacetic acid). Electrophoresis was performed in TAE buffer for 60mins at 120V. To estimate the size of DNA fragments, samples was electrophoresed alongside either a 100bp or 1kb marker (NEB), as appropriate for the analysis.

3.2.2 Plasmid and Genomic DNA Preparation

Plasmid DNA preparation was performed with a Qiagen MiniPrep kit according to the manufacturers protocol. Genomic DNA preparation was performed using a Qiagen Blood and Tissue kit.

3.2.3 Restriction Digests

Analytical and preparative restriction digests were performed for 1hr at 37ºC in a thermostatic water bath with a concentration of restriction enzyme (New England Biolabs) according to the manufacturer's guideline of 1 unit per 1µg of DNA.

3.2.4 Restriction Cloning

Vectors and inserts were digested as above and agarose electrophoresis of the entire reaction products performed. Inserts and vector backbones were excised from the gel and purified using a Qiagen Qiaquick kit. Purified digested vectors and inserts and were mixed in a range of ratios

between 1:3 and 1:1 in DNA Ligase Buffer (New England Biolabs) and incubated for 2hrs at room temperature with 1 unit of T4 DNA Ligase (New England Biolabs). Transformation of *E. coli* 10β with ligation products was performed as in Section 3.1.4 and plasmid stocks subsequently generated as in Section 3.2.2. Cloning of the insert was analysed by analytical digestion as in Section 3.2.3.

3.2.5 Polymerase Chain Reaction (PCR)

Polymerase chain reactions were assembled as in Table 3.2.1. Cycling was performed with a PTC-200 DNA Engine (Bio-Rad) according to Table 3.2.2. Optimum annealing temperatures, detailed in Table 3.2.3, were determined means of an annealing temperature gradient of 5ºC either side of the calculated optima provided by Eurofins MWG. Extension time was calculated based on the manufacturer's guidelines of 1min per 1.5kb of amplified product. Where colony PCR was performed, template DNA was 1µl of a colony emulsion generated by resuspending a single colony in 100µL molecular biology grade water.

Volume (µl) per reaction	**Component**
12.5	Failsafe Buffer E (Epicentre Biotech)
0.5	(1.75 Units) Expand HiFi DNA Polymerase blend (Roche)
1	1ul Forward primer (1pM final concentration)
1	1ul Reverse primer (1pM final concentration)
1	1ul Template DNA
9	9ul Molecular biology quality water

Table 3.2.1 – Assembly of Polymerase Chain Reactions (PCR).

Step	**Temperature (ºC)**	**Time**	
Denaturation	94	3min	
Denaturation	94	15s	x 25 cycles
Annealing	See Table XX	15s	
Extension	72	1min per 1.5kb of amplicon	
Extension	74	3min	

Table 3.2.2 – PCR Cycling Conditions.

3.2.6 Nucelotide Sequencing

DNA sequencing, by the cycle sequencing/chain termination method, was performed by Eurofins MWG Operon using an ABI 3730XL sequencer (Applied Biosystems).

3.2.7 Oligonucelotide PCR Primers

Synthesis of oligonucleotides, by the phosphoramidite method, was performed by Eurofins MWG Operon. Primer sequences and properties are listed in Table 3.2.3.

Name	**Sequence (5'-3')**	**Annealing temperature (°C)**
CD2718FlankF	TTCACTCAAAACCTTCACTCC	55.9
CD2718FlankR	TCGATTCCTATCACCAGCTC	
CD2718 Intern F	AACCTTGAATGGAAAAGAAT	50.0
CD2718 Intern R	GCATCATCAAATTCATAGGT	
rpoB F	CAATAGAGACTCCAGAGGGAC	59.8
rpoB R	AGGAACCATTTCAACAGCAC	
CD2718C26A F	GTTACGCTGTCTACTGCTACTTACGAATTCG	65
CD2718C26A F	CGAATTCGTAAGTAGCAGTAGACAGCGTAAC	
CD0386 pET28F	GCGAGAGAGAGCATATGATGAAGCTTGAGAG CAAACAATA	57.3
CD0386 pET28R	GAGGACTCTCGAGTTACTTCACACCTTCGAT ATCGCCACG	
ErmF	ACGCGTTATATTGATAAAAATAATAATA GTGGG	50.0
ErmR	ACGCGTGCGACTCATAGAATTATTTCCT CCCG	

Table 3.2.3 – Sequence and Properties of PCR Primers.

3.2.8 Gene Synthesis

Gene synthesis was performed by DNA 2.0 (Menlo Park, California USA), with the exception of the *srtB* gene, which was synthesised by Entelechon Gmbh.

3.3 General Protein Methods

3.3.1 Analysis of Proteins by Polyacrylamide Gel Electrophoresis

Protein samples of 30µl were mixed with 10µl NuPAGE loading buffer (Invitrogen) and heated at 90°C for 5mins. After cooling, 10µl of the prepared sample was loaded into the wells of a 4-12% bis-tris acrylamide gel measuring 80mm x 8mm x 10mm (Invitrogen). Electrophoresis was performed for 35mins within an X-cell Surelock chamber, in a buffer of 50 mM MES pH7.3, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA and at a voltage of 200V. For estimation of protein molecular weights, all samples were run alongside SeeBlue Plus2 pre-stained standards (Invitrogen).

Gels were coomassie stained by placing in 50ml of SimplyBlue stain (Invitrogen), heating for 2mins in a 700W microwave oven then incubating for 15mins with rocking agitation. Stain was decanted from the tray and replaced with an equal volume of distilled water, microwaved for 2mins then incubated for 30mins with rocking agitation. Gels were rinsed once more in 50ml distilled water overnight to remove all stain. All gels were documented and analysed using a Bio-Rad GS-800 densitometer and Quantity One analysis software (Bio-Rad).

3.3.2 Protein Assays

For routine determination of protein concentration, absorbance at 280nm was measured using a Nanovue small volume spectrophotometer (GE healthcare) and calculation of concentration based on molar extinction coefficient as provided by analysis of the protein sequence using the Protperam utility (http://web.expasy.org/protparam/) and as detailed in Appendix 1.

Where noted in methods, a bicinchoninic acid BCA (Thermo) assay was utilised according to the manufacturer's instructions, with a range of Bovine Serum Albumin (BSA) standards between 20-2000µg/ml and ensuring dilution of the sample to fall within this range.

3.3.3 Western Blotting

For western blotting, SDS-PAGE was performed as above. Gels were washed briefly with distilled water, overlaid with a nitrocellulose membrane of 0.45µm pore size (Invitrogen) and placed within

filter paper sandwich in an Xcell II blot module (Invitrogen). Blotting was performed for 60mins at 30V in NuPAGE transfer buffer (Invitrogen).

Membrane was washed once in TBS-T (50mM Tris pH 7.5, 150mM NaCl, 1% Tween 20) and blocked by incubation at 4ºC overnight in 50ml of blocking buffer (TBS-T, 5% skimmed milk powder). Primary antibody was diluted to the appropriate concentration in 50ml blocking buffer and incubated at room temperature for 90mins with rocking agitation and then thrice washed for 5mins in 50ml TBS-T before addition of an alkaline phosphatase conjugated secondary antibody diluted to the appropriate concentration in 50ml blocking buffer and then thrice washed for 5mins in 50ml TBS-T before addition of nitro-blue tetrazolium/5-bromo-4-chloro-3'-indolyphosphate (NBT/BCIP) reagent (Pierce). Development was halted after 10mins by extensive washing with distilled water.

3.3.4 Recombinant Protein Expression Tests

To select clones and conditions for optimal expression of recombinant proteins, 10ml of TB was inoculated with a single half colony picked from a plate of fresh transformants and incubated at 37°C with orbital agitation at 180rpm. Upon the culture growing to an optical density of 0.6 at 600nm, Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a concentration of 1mM. After incubation for a further 3hrs, aliquots of 1ml were taken and sonicated at for 30s. Lysates were centrifuged at 10,000g for 10mins. The supernatant, containing soluble protein, was decanted while the pellet containing insoluble protein was resuspended in 1ml of Phosphate Buffered Saline (PBS) and sonicated for 30s. 10µl of each fraction was analysed according to Section 3.3.1. For archiving of selected clones with maximal expression, the remaining half colony was picked from the plate, emulsified in the cryoprotectant fluid of a Cryobank bead stock tube (Copan Diagnostics), the fluid removed by pipetting and the vial stored at -80°C.

3.3.5 Recombinant Protein Expression

Archived bead stocks of *E. coli* BL21 were retrieved from storage at -80°C and a single bead used to inoculate 50ml of TB supplemented with the appropriate antibiotic as detailed in Section 3.1.4. After overnight incubation at 37°C with orbital shaking at 180rpm, a 5% subculture was made into 1L of fresh broth and incubation continued while optical density at 600nm was monitored at hourly intervals. Upon the culture growing to an optical density of 0.6 at 600nm, IPTG was added to a concentration of 1mM and the temperature of incubation lowered to 16°C for a further 16 hours.

Cell pellets were recovered by centrifugation for 30mins at 3,500g in a Sorvall RC3BP centrifuge with a H6000A rotor and resuspended 10% w/v in the chromatography 'Buffer A' as detailed. Cells were lysed by sonication on ice for 10 x 30s intervals and cell debris removed by centrifugation for 30mins at 40,000g in a Sorvall RC5C with an SS-34 rotor. Supernatants were filtered with a 0.22µm cellulose syringe filter (Sartorius) and processed as detailed in Section 4.2. All procedures were performed under refrigeration at 4 ºC where possible or on ice.

Chapter 4 – Purification of Recombinant Proteins

4.1 Introduction

This chapter describes the expression and purification of four recombinant proteins originating in *C. difficile* which are utilised for various purposes throughout the remainder of the thesis. The proteins are briefly described below in Table 4.1, along with references to their use later in the work.

Name	Description/Function	Purpose(s)	Section Reference
SrtB	Putative *C. difficile* Sortase	Raising of antiserum Crystallisation Biochemical Assays	Section 5.2 Section 6.2 Section 8.2
SrtB C226A	Site directed mutant of SrtB, Cys to Ala mutation at position 226	Crystallisation Biochemical Assays	Section 6.2 Section 8.2
CD0386	A putative wall-sorted protein with homology to the *S. aureus* Cna collagen binding protein and the RrrA pilin of *Streptococcus pneumoniae*	Raising of Antiserum	Section 5.2
CD0386N	N-terminal Truncation mutant of CD0386	Crystallisation	Section 7.2

Table 4.1 – Proteins Expressed and Purified in this Chapter – Section references correspond to to the methods section where the protein or its antiserum is predominantly used.

4.2 – Methods

4.2.1 Purification of *C. difficile* SrtB

A codon-optimised gene encoding *C. difficile* SrtB, devoid of its predicted signal peptide (See Appendix A for sequence), was synthesised by Entelechon Gmbh and supplied within the plasmid pEXP1. Protein was expressed as in Section 3.3.5, and the following buffers were prepared:

Buffer A – 50mM HEPES pH 7.5, 500mM NaCl, 10mM imidazole
Buffer B – 50mM HEPES pH 7.5, 500mM NaCl, 500mM imidazole
Buffer C – 50mM HEPES pH 7.5, 1M $(NH_4)_2SO_4$
Buffer D – 50mM HEPES pH 7.5

A XK16 chromatography column (GE Healthcare) was packed with 20ml of Chelating Sepharose (GE Healthcare), charged with nickel as per the manufacturer's instructions and equilibrated with 10 column volumes (CV) of Buffer A. Lysate was applied at a rate of 2ml/min, following which the column was washed with Buffer A at 2ml/min until the measured absorbance at 280nm had returned to the baseline level. Flow rate was reduced to 1ml/min and a buffer gradient initiated from 100% Buffer A to 100% Buffer B over a period of 40mins, during which time 5ml fractions were collected.

Fractions corresponding to absorbance peaks in the chromatograph were analysed by SDS-PAGE, and those containing the protein of interest were pooled and dialysed overnight in 5L of Buffer C at 4ºC. A 5ml HiTrap Butyl HP column (GE healthcare) was equilibrated with 10CV of Buffer C after which the dialysed eluate from the nickel affinity step was applied at a rate of 1ml/min, following which Buffer D was applied at the same rate until the measured absorbance at 280nm had returned to the baseline level.

SrtB was dialysed extensively in a storage buffer consisting of 25mM HEPES, 150mM NaCl, concentrated to a target concentration of 10mg/ml using a 20ml Vivaspin 10kDa centrifugal filtration unit and stored at -80°C .

4.2.2 Site Directed Mutagenesis of pEXP1-SrtB to generate pEXP1-SrtB C226A

Mutagenesis of *srtB* to replace the 'TGT' codon encoding cystiene 226 with 'GCT', encoding an alanine residue, was performed using a method based on the technique commercially known as 'Quickchange PCR' (Weiner and Costa, 1994). The plasmid pEXP1-*srtB* was amplified in its entirety by PCR with primers complementary to the region for mutation, but incorporating the

desired nucleotide changes (Table 3.2.3, CD2718C226A F and CD2718C226A R). Restriction digestion of PCR products was performed with the methylation sensitive restriction enzyme DpnI , in order to cleave the methylated template plasmid while leaving the newly synthesised PCR products intact. After analysis by agarose gel electrophoresis, NEB 10β cells were transformed with the DnpI digestion products. Plasmid was prepared from transformants and a restriction digest with RsaI performed to screen for mutants by the loss of an RsaI site which overlapped the codon encoding Cys226. Mutants were confirmed by sequencing PCR fragments amplified with primers designed to anneal within the T7 promoter and terminator.

4.2.3 Purification of *C. difficile* SrtB C226A

Expression and purification was performed as for SrtB.

4.2.4 Purification of *C. difficile* CD0386 (C-terminal His Tag)

A gene encoding CD0386 devoid of its signal peptide and incorporating a C-terminal thrombin-cleavable hexahistidine tag was synthesised by DNA 2.0 and supplied within the plasmid pJ414 (See Appendix A for sequences). Protein was expressed as in Section 3.3.5, and the following buffers were prepared:

Buffer A – 50mM HEPES pH 7.5, 500mM NaCl, 10mM imidazole
Buffer B – 50mM HEPES pH 7.5, 500mM NaCl, 500mM imidazole

An XK16 chromatography column (GE Healthcare) was packed with 20ml of chelating sepharose (GE Healthcare), charged with nickel according to the maufacturers instructions and equilibrated with 10 column volumes (CV) of Buffer A. Lysate was applied at a rate of 2ml min, following which the column was washed with Buffer A at the same rate until the measured absorbance at 280nm had returned to the baseline level. Flow rate was reduced and a buffer gradient initiated from 100% Buffer A to 100% Buffer B over a period of 40mins, during which time 5ml fractions were collected.

4.2.5 Cloning and Purification of CD0386 (N-terminal His Tag)

To express CD0386 with an N-terminal affinity tag, restriction cloning via HindIII/KpnI was performed according to Section 3.2.4 in order to clone CD0386 into pTAC-MAT1 (Invitrogen) which encodes a metal affinity fusion tag of the sequence ‘MAHNHRKH’. Correct cloning of the gene was confirmed by analytical restriction digest with HindIII/KpnI as in Section 3.2.3. Protein was expressed as in Section 3.3.5 and the following buffers prepared:

Buffer A – 50 mM HEPES pH7.5, 500mM NaCl, 10mM imidazole
Buffer B – 50mM HEPES pH 7.5, 500mM NaCl, 500mM imidazole
Buffer C – 20mM Bis-Tris pH 5.5, 180mM NaCl
Buffer D – 20mM Bis-Tris pH 5.5, 500mM NaCl

A XK16 chromatography column (GE Healthcare) was packed with 20ml of chelating sepharose (GE Healthcare), charged with nickel according to the maufacturers instructions and equilibrated with 10 column volumes (CV) of Buffer A. Lysate was applied at a rate of 2ml min, following which the column was washed with Buffer A at the same rate until the measured absorbance at 280nm had returned to the baseline level. Flow rate was reduced and a buffer gradient initiated from 100% Buffer A to 100% Buffer B over a period of 40mins, during which time 5ml fractions were collected. Fractions corresponding to absorbance peaks in the chromatograph were analysed by SDS-PAGE, and those containing the protein of interest were pooled and dialysed overnight in 5L of Buffer C at 4ºC.

A MonoQ 5/50 anion exchange column (GE Healthcare) was equilibrated with 5CV of Buffer C and the pooled dialysed fractions applied at a rate of 1 ml/min. The column was washed at the same rate with buffer C until absorbance at 280nm returned to the baseline level. A gradient to 25% Buffer D over 30mins was initated and 1ml fractions collected.

CD0386 was dialysed extensively in a storage buffer consisting 25mM HEPES, 150mM NaCl, concentrated to a target concentration of 10mg/ml using a 20ml Vivaspin 10kDa centrifugal filtration unit and stored at -80°C

4.2.6 Cloning and Purification of an N-terminal Fragment of CD0386 (CD0386N)

For crystallographic studies, an N-terminal fragment of CD0386 was selected, encompassing a domain identified by Phyre2 as having predicted secondary structure homology to the RrgA pilin of *Streptococcus pneumoniae* (Chapter 7 describes this fragment in detail, see Appendix A for sequence). The fragment was amplified from pJ414-CD0386 with primers incorporating restriction sites to allowing cloning into pET28a via NdeI and XhoI (Table 3.2.3, CD0386 pET28F and pET28R). Protein was expressed as in Section 3.3.5. For purification, the following buffers were prepared:

Buffer A – 50mM HEPES pH 7.5, 500mM NaCl, 10mM imidazole
Buffer B – 50mM HEPES pH 7.5, 500mM NaCl, 500mM imidazole

A XK16 chromatography column (GE Healthcare) was packed with 20ml of Chelating Sepharose (GE Healthcare), charged with nickel as per the manufacturer's instructions and equilibrated with 10 column volumes of Buffer A. Lysate was applied at a rate of 2ml min, following which the column was washed with Buffer A at the same rate until the measured absorbance at 280nm had returned to the baseline level. Flow rate was reduced to 1ml/min and a buffer gradient initiated from 100% Buffer A to 100% Buffer B over a period of 40mins, during which time 5ml fractions were collected.

After analysis of the nickel affinity purification by SDS-PAGE, fractions containing CD0386N were concentrated to a volume of 2ml and applied to a 16/60 Superdex 200 at a rate of 0.5ml/min with a running buffer of 50mM HEPES pH7.5, 250mM NaCl. After SDS-PAGE analysis, peak fractions were collected and concentrated to a target concentration of 10mg/ml using a 20ml Vivaspin 10kDa centrifugal filtration unit (Sartorius AG).

4.2.7 Production and Titration of Rabbit Antiserum against SrtB and CD0386

Production of rabbit antiserum was performed by Covalab S.A.S. A 50µg dose of either SrtB or CD0386 was mixed with freund's incomplete adjuvant and injected subcutaneusly into duplicate New Zealand White rabbits. Immunisations were performed on days 0, 21 and 42. A terminal bleed was performed on day 53.

To determine antibody titre, an Enzyme Linked Immunosorbent Assay (ELISA) was performed. The antigen to which the serum was raised was diluted to 5µg/ml in Phosphate Buffered Saline (PBS) and 100µl added to rows 1-6 of a 96 well ELISA plate (Corning) while 100µl of PBS was applied to the remaining wells. After overnight incubation at 4ºC, the plate was washed thrice with PBS + 0.1% Tween 20 (PBS-T) and blocked by incubation with blocking buffer (PBS-T + 5% foetal calf serum) for 2hrs at 37ºC. The serum was diluted 1 in 2 in blocking buffer and 200µl of this dilution applied to wells in column 1. 120µL of blocking buffer was placed in the remaining wells in row A-C, and a 60µL serial dilution was performed across the plate resulting in 1 in 3 dilution curve. The plate was incubated at 37°C for 2hrs followed by washing thrice with PBS-T and the application of the secondary antibody (Sigma goat anti-rabbit HRP conjugate) at a dilution of 1:10,000 in blocking buffer. Plate was incubated at 37°C for 2 hours followed by thrice washing with PBS-T and the application 100µL of 3,3',5,5'-Tetramethylbenzidine (TMB). 100µL of TMB stop reagent was added after 5mins. Absorbance at 450nm was measured using a plate spectrophotometer (Tecan) using 620nm as a reference frequency.

4.2.8 – Analytical Gel Filtration of SrtB and SrtB C226A

A Superdex 200 analytical gel filtration column (GE Healthcare) was washed with one column volume of filter sterilised water followed by one column volume of filter sterilised phosphate buffered saline (PBS) at a flow rate of 0.3ml/min. 40µg of SrtB or SrtB C226A was injected onto the column in a total volume of 200µl and isocratic elution with PBS continued at 0.3ml/min while absorbance was recorded at 280nm. Molecular weights estimation was performed by reference to a standard curve prepared by analysis of Blue Dextran 2000, Thyroglobulin, Ferritin, Aldolase, Conalbumin and Ovalbumin.

4.3 Results

4.3.1 Purification of SrtB

C. difficile SrtB was expressed in batch culture of *E. coli* BL21 from plasmid pEXP1-SrtB, which encodes a fusion protein comprising of an enterokinase cleavable hexahistidine tag. The purification method, detailed in Section 3.4.1, consisted of a nickel affinity chromatography followed by hydrophobic interaction chromatography (see Figure 4.3.1). Overall yield was typically 5mg per litre of culture. Mass spectroscopic analysis of the purified protein revealed a mass within 1 dalton of the theoretical mass of the protein (see Figure 4.3.3).

4.3.2 Mutagenesis and Purification of SrtB C226A

C. difficile SrtB C226A was expressed in batch culture of *E. coli* BL21 from plasmid pEXP1-SrtB C226A, generated by site directed mutagenesis of pEXP1-SrtB and hence encoding the same enterokinase cleavable hexahistidine tag. Purification methodology and typical yield was as for SrtB (See Figure 4.3.2). Mass spectroscopic analysis of the purified protein revealed a mass within 1 dalton of the theoretical mass of the protein and with a loss of 32 Da relative to SrtB as expected due to the C226A mutation (see Figure 4.3.4)

4.3.3 Attempted Removal of Hexahistidine tag from *C. difficile* SrtB

C.difficle SrtB was incubated overnight at 4ºC with a varying range of concentration of enterokinase in an attempt to remove the hexahistidine tag. A multiple band pattern consisting of apparent masses of 28 , 23, 22, 20, 6, 5 and 4 kDa was observed in all cases (see Figure 4.3.5).

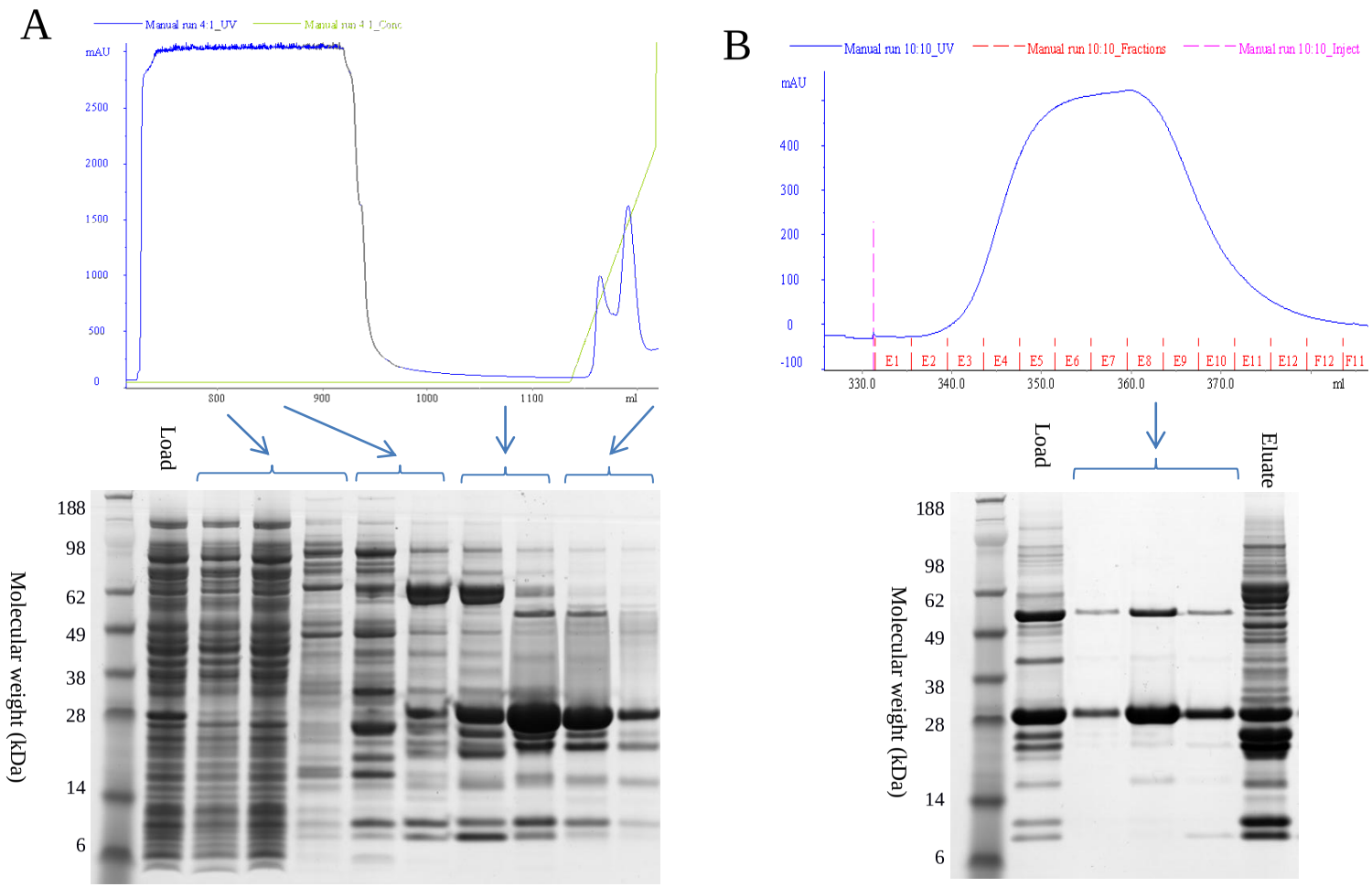


Figure 4.3.1. Chromatography and SDS-PAGE Analysis of SrtB Purification. A) Chromatogram illustrating binding and imidazole gradient elution of SrtB from a Ni^{2+} charged chelating sepharose and (below) SDS-PAGE of the the resulting fractions B) Chromatogram illustrating removal of contaminants from Ni^{2+} affinity purified SrtB by flowing through a 5ml butyl sepharose column in the presence of 1M ammonium sulphate and (below) SDS-PAGE analysis demonstrating effective removal of contaminants.

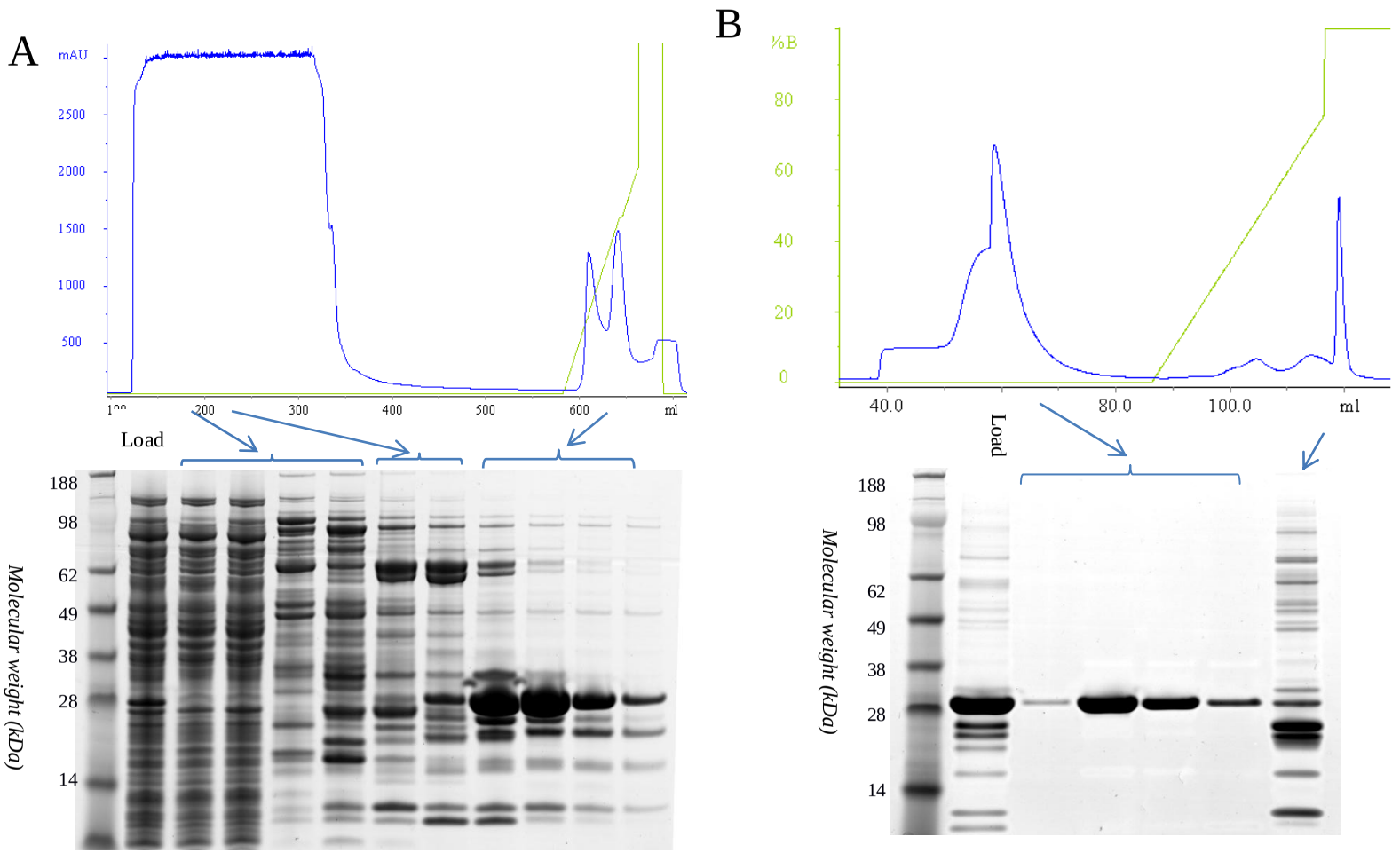


Figure 4.3.2. Chromatography and SDS-PAGE Analysis of SrtB C226A Purification. A) Chromatogram illustrating binding and imidazole gradient elution of SrtB C226A from a Ni^{2+} charged chelating sepharose and (below) SDS-PAGE of the the resulting fractions B) Chromatogram illustrating removal of contaminants from Ni^{2+} affinity purified SrtB C226A by flowing through a 5ml butyl sepharose column in the presence of 1M ammonium sulphate and (below) SDS-APGE analysis demonstrating binding of contaminants to the column.

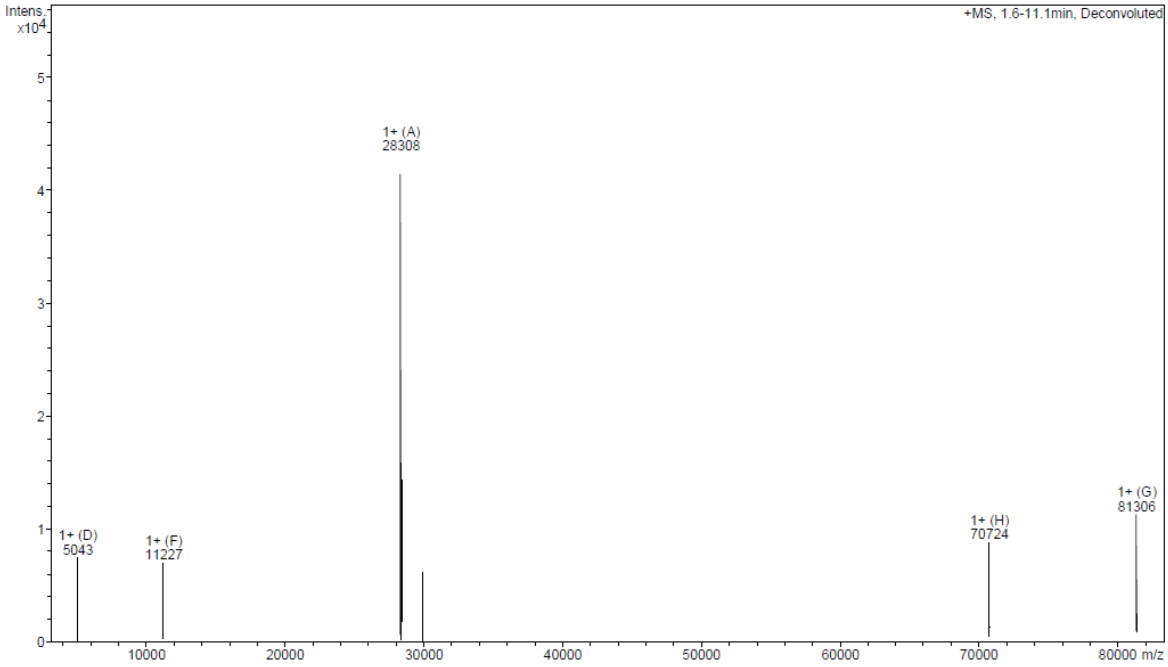


Figure 4.3.3 – Mass Spectroscopic Analysis of Purified SrtB. The major peak of 28308 Da is in agreement with the theoretical mass of SrtB of 28307.6 Da. Minor contaminants are observed at 5, 11, 70 and 81kDa.

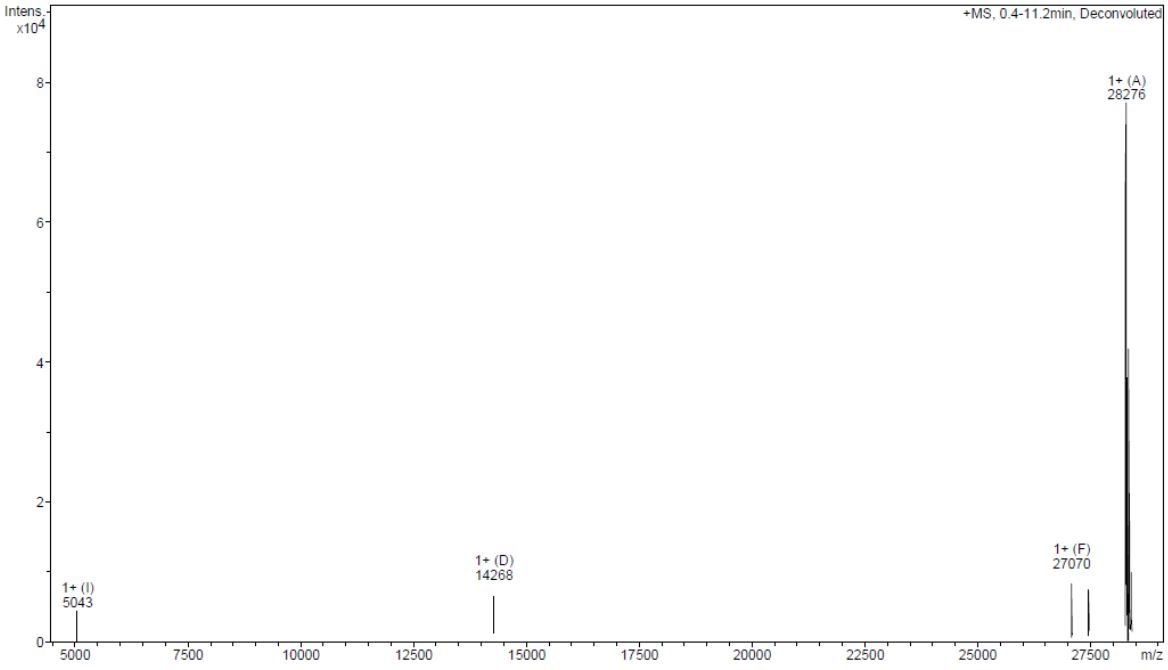


Figure 4.3.4 – Mass Spectroscopic Analysis of Purified SrtB C226A. The major peak of 28276 Da is in agreement with the theoretical mass of SrtB C226A of 28275.5 Da. Minor contaminants are observed at 5, 14 and 27 kDa.

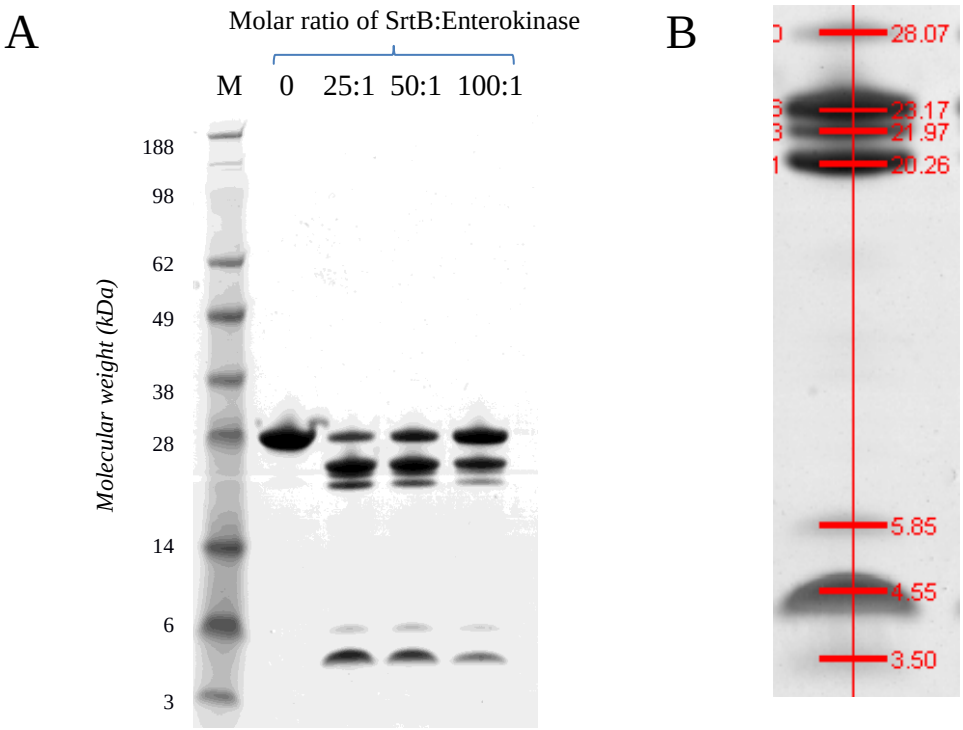


Figure 4.3.5. Attempted Removal of the Polyhistidine Tag of SrtB. A) SDS-PAGE of cleavage products following incubation of SrtB with enterokinase overnight at 4°C B) Enlargement of lane 2 with estimated molecular weights as determined by a standard curve.

4.3.4 Purification of CD0386 (C-terminal Polyhistidine tag)

Pilot nickel affinity purification (see Figure 4.3.6) of CD0386 expressed in *E. coli* from plasmid pJ414 with the inclusion of a C-terminal hexahistidine tag resulted in poor yield and purity which was not improved even after extensive method optimisation.

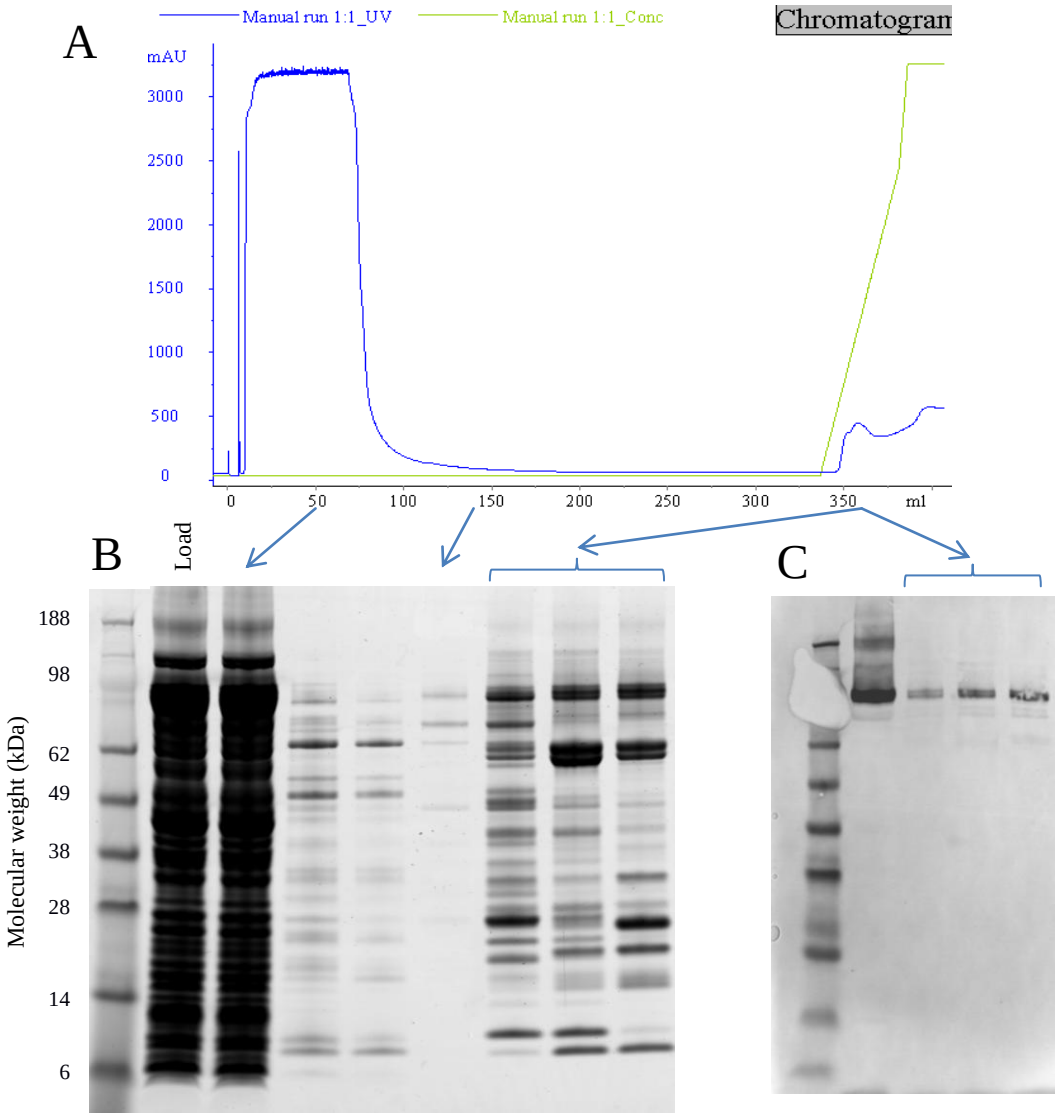


Figure 4.3.6 – Purification of CD0386 (C-terminal polyhistidine tag). Chromatogram illustrating binding and imidazole gradient elution of C-terminally His-tagged CD0386 from a Ni^{2+} charged chelating sepharose. B) SDS-PAGE of the resulting fractions C) Western blotting and probing of selected fractions with an anti-histidine antibody.

4.3.5 Cloning and Purification of CD0386 (N-terminal metal affinity tag)

In order to improve yield and purity of CD0386, the gene was restriction cloned into pTAC-MAT1, a vector encoding a N-terminal metal affinity fusion tag. Expression of this construct and purification by nickel affinity resulted in much improved yield and purity (Figure 4.3.7 A). Lower molecular weight contaminants were substantially removed by a downstream ion exchange chromatography (Figure 4.3.7 B).

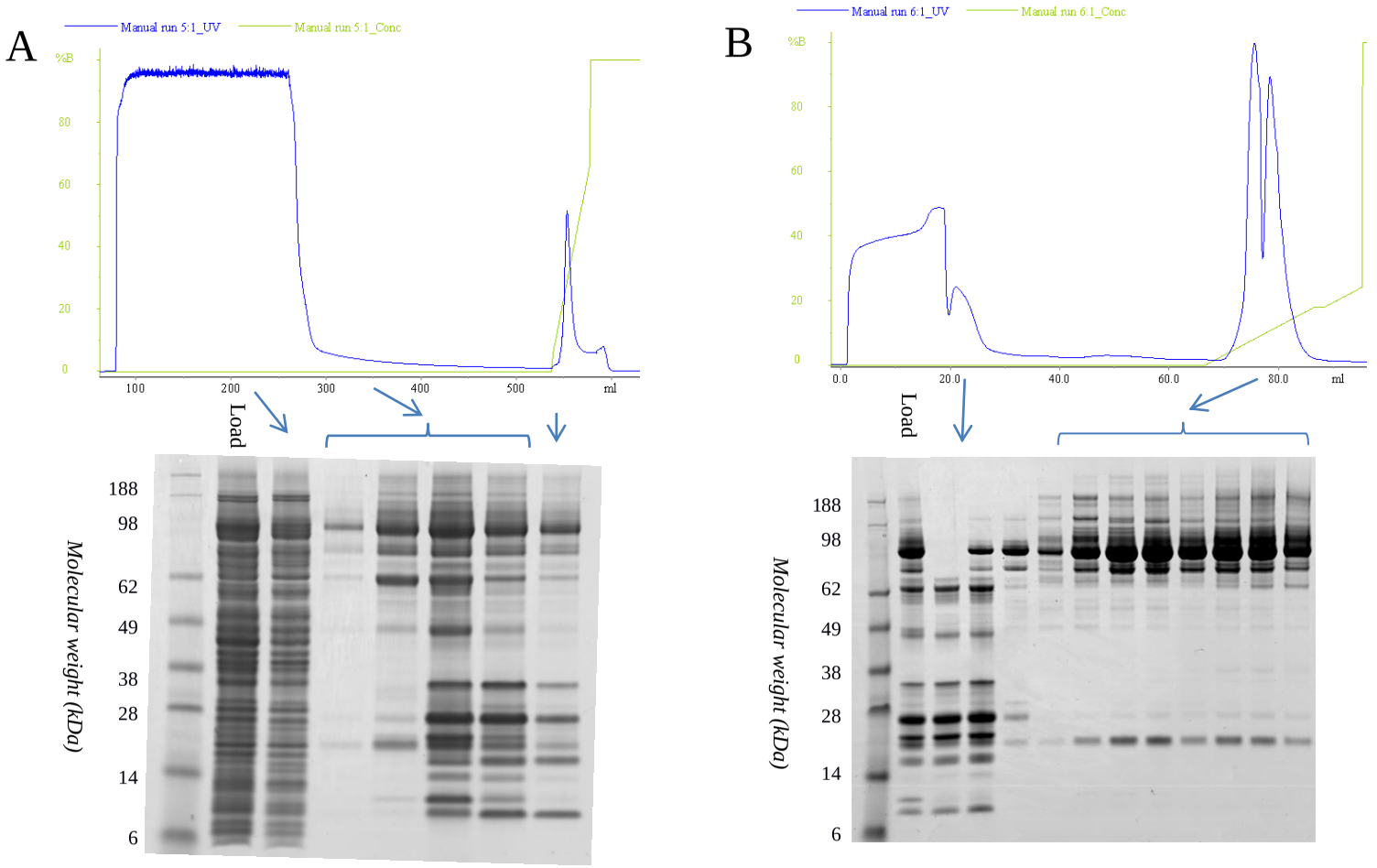


Figure 4.3.7 Chromatography and SDS-PAGE Analysis CD0386 (N-terminal His Tag) Purification. A) Chromatogram illustrating binding and imidazole gradient elution of CD0386 from a Ni^{2+} charged chelating sepharose and (below) SDS-PAGE of the the resulting fractions B) Chromatogram illustrating removal of contaminants from Ni^{2+} affinity purified CD0386 by MonoQ ion exchange chromatography and (below) SDS-PAGE of the resulting fractions.

4.3.6 Cloning and Purification of CD0386N

A gene fragment encoding N-terminal domain of CD0386 (See Appendix A for sequence information) was cloned into pET28a in frame with an N-terminal thrombin cleavable hexahistidine tag and expressed in *E. coli* BL21. Purification consisted of nickel affinity chromatography (Figure 4.3.9 A), thrombin cleavage of the tag and recovery by nickel affinity chromatography followed by size exclusion chromatography to remove contaminating high molecular weight aggregates (Figure 4.3.9 B). Mass spectroscopic analysis of the purified protein confirms a mass of within 3Da of the theoretical mass (Figure 4.3.8).

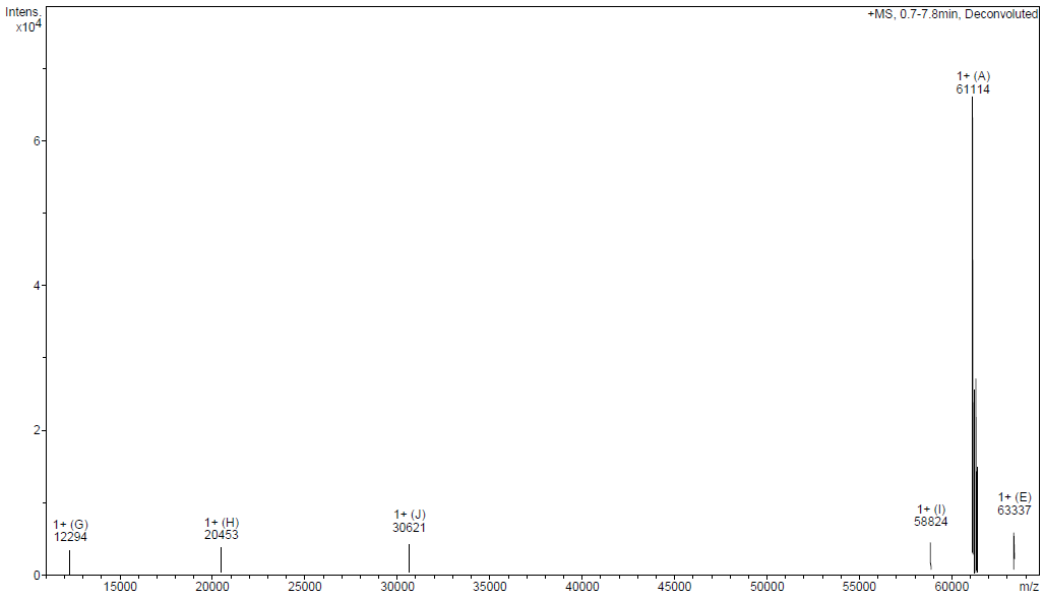


Figure 4.3.8 – Mass Spectroscopic analysis of purified CD0386N. The major peak of 61114 Da is in agreement with the theoretical mass of thrombin-cleaved CD0386N of 61111 Da. Minor contaminants are present at 12, 20, 30, 59 and 63kDa. No peak is observed for residual uncleaved CD0386 (mass 62630.9).

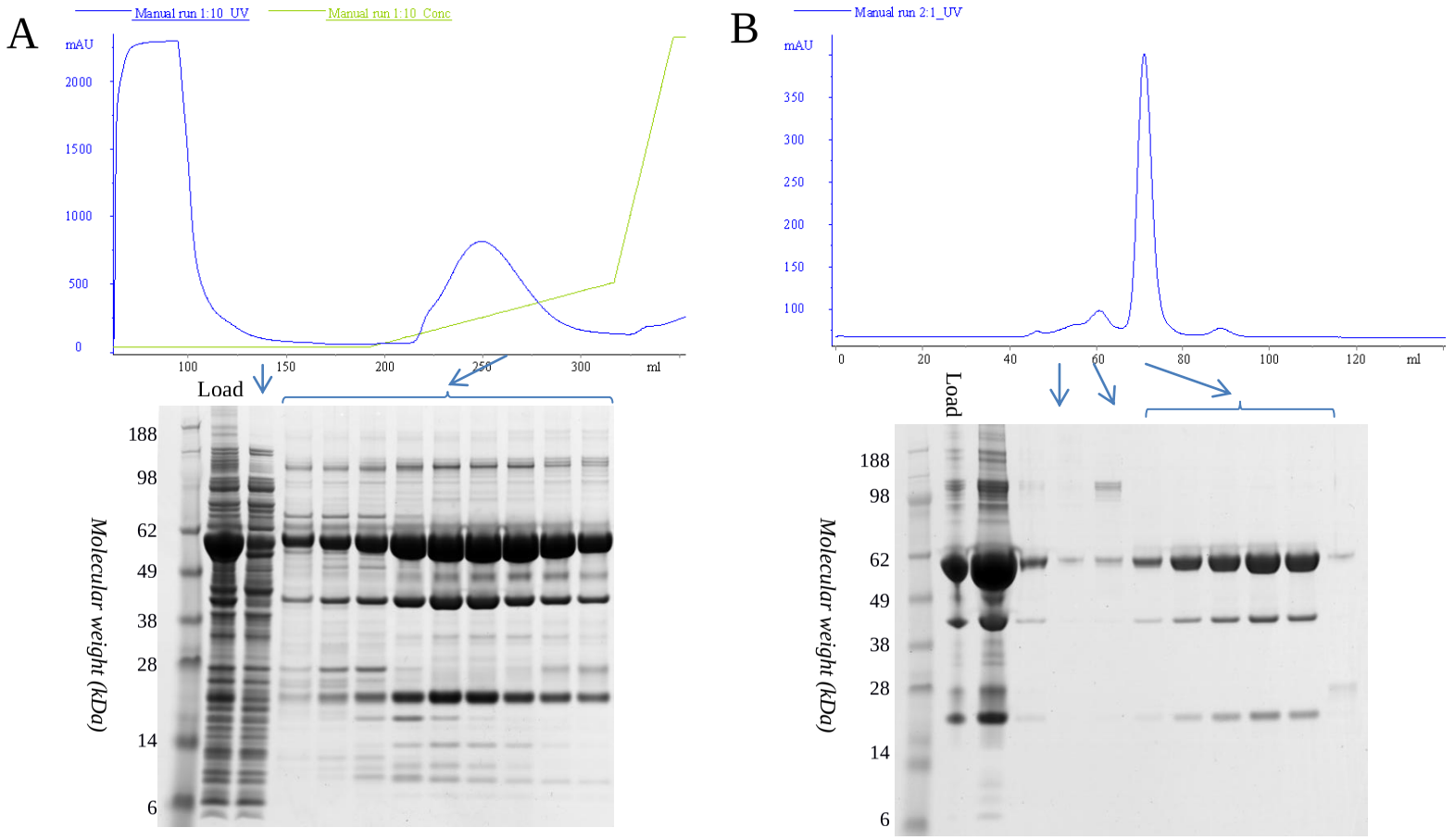


Figure 4.3.9 - Chromatography and SDS-PAGE Analysis of CD0386N Purification. A) Chromatogram illustrating binding and imidazole gradient elution of CD0386N from a Ni^{2+} charged chelating sepharose and (below) SDS-PAGE of the the resulting fractions B) Chromatogram illustrating removal of contaminants from Ni^{2+} affinity purified and thrombin-cleaved CD0386N by size exclusion chromatography using a Superdex 200 column and (below) SDS-PAGE of the resulting fractions.

4.3.7 Production and Titration of Rabbit Antiserum Against CD2718 and CD0386

Rabbits were immunised in duplicate and their serum titrated by ELISA on plates coated with the appropriate antigen (Figure 4.3.10). For both antigens, titres of ~10^{-7} were obtained with little difference between individual rabbits. Sacrifice of rabbits on Day 53 produced a terminal bleed of 40-50ml per animal.

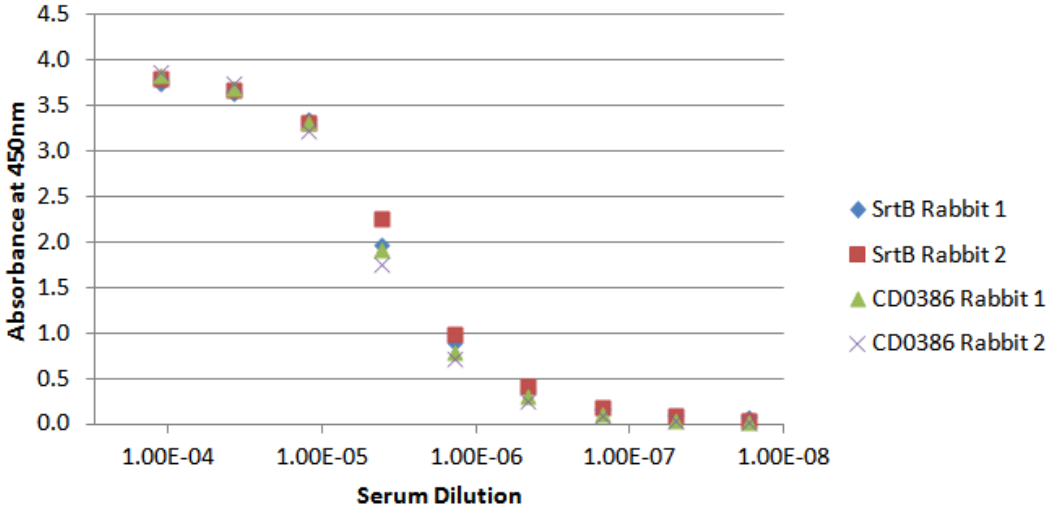


Figure 4.3.10 – Titration of anti-SrtB and anti-CD0386 Serum. Duplicate rabbits were immunised with recombinant protein. Terminal bleeds on Day 53 were analysed by ELISA as in Section 4.2.7.

4.3.8 – Analytical Gel Filtration of SrtB and SrtB C226A

Analytical Gel Filtration of SrtB and SrtB C226A was performed as described in Section 4.2.8. Chromatographs are shown in Figure 4.3.11

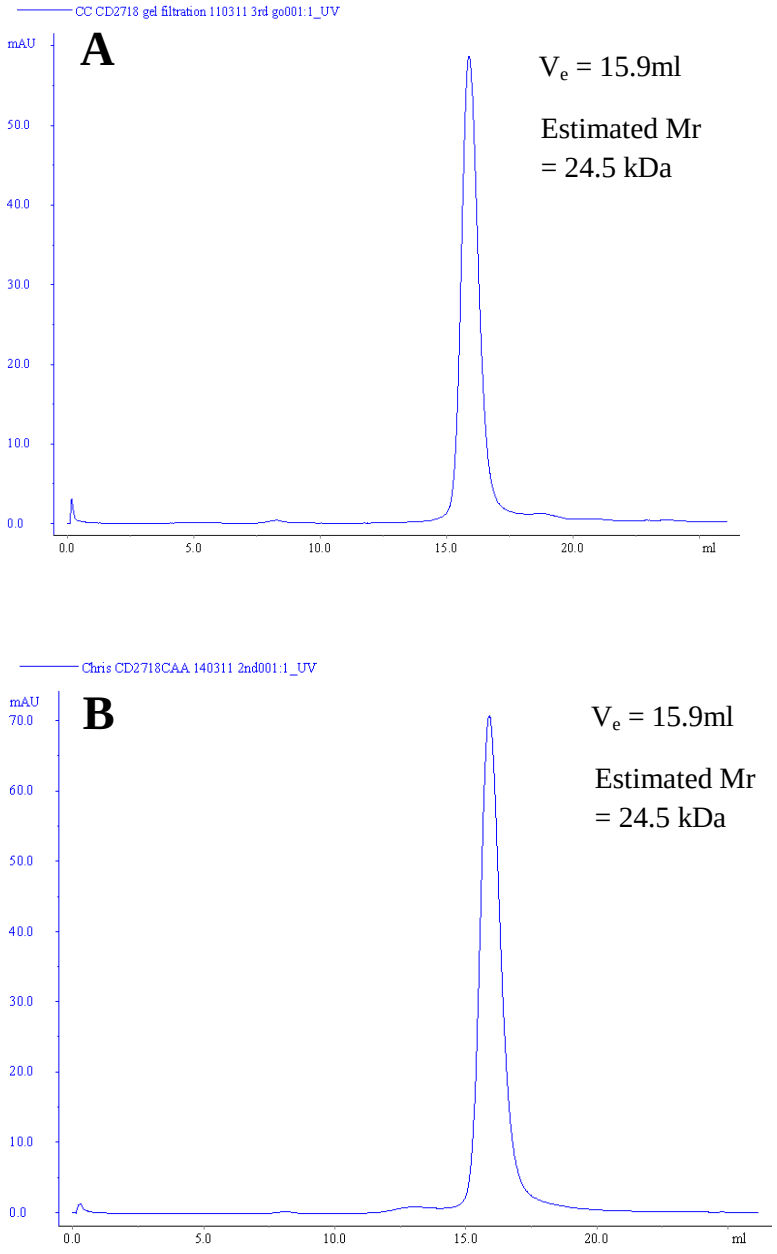


Figure 4.3.11 - Analytical Gel Filtration of SrtB (A) and SrtB C226A (B). Estimated mass was determined by reference to a standard curve generated by analysis of six molecular weight standards.

4.4 Discussion on Purification of Recombinant Proteins

4.4.1 Purification of *C. difficile* SrtB and SrtB C226A

A good yield of SrtB was obtained following nickel affinity purification, with in excess of 20mg typically purified from a 4 litre batch culture. A negative chromatography step comprising a butyl sepharose hydrophobic interaction column increased the purity of nickel affinity purified SrtB to approaching 100% as determined by SDS-PAGE. Purification of the SrtB C226A mutant was essentially identical in both method and results with the exception of dimerisation as described below. Mass spectroscopic analysis indicates that both proteins are of the expected mass with the C226A substitution clearly apparent with mass difference of 32Da.

Some dimerisation was observed in non-reducing SDS-PAGE analysis of SrtB (see Figure 4.3.1 B, ~56kDa band). This dimerisation was not observed in reducing SDS-PAGE or in SDS-PAGE of the C226A mutant. As SrtB contains no cysteine other than Cys226, this dimerisation must be due to disulphide bonding between Cys226 residues of SrtB active site. Gel filtration experiments provide an estimated monodisperse mass of 24.5kDa, confirming that the protein is monomeric rather than dimeric in solution. It appears, therefore, that disulphide formation is an artefact of preparation for SDS-PAGE, perhaps due to exposure of the buried active site sulphydryl during heating or denaturation.

4.4.2 Removal of Polyhistidine tag from SrtB

The plasmid from which SrtB was expressed, pEXP1, encodes a spacer peptide between the cloned gene and the polyhistidine tag which contains a canonical enterokinase cleavage site 'DDDDK' . However, despite its specificity for a pentapeptide site, enterokinase is known to cleave at non-canonical sites (Shahravan *et al.*, 2008). When removal of the tag from SrtB was attempted, a multiple band pattern was observed indicating that enterokinase cleaves SrtB at multiple sites. Shahavran *et al* describe methods to improve the specificity of enterokinase cleavage by denaturation of the tagged protein but this is not a desirable methodology for purification of an active enzyme. If it was deemed strictly necessary to remove the tag from this protein, cloning into a vector with a tag cleavable with a different protease (eg. Thrombin or Factor Xa) could be performed.

4.4.3 Purification of CD0386

Following nickel affinity chromatography, neither yield nor purity of CD0386 was satisfactory when expressed with an C-terminal hexahistidine tag. Blotting with an anti-his antibody did,

however, indicate expression of a protein of the expected mass. Poor affinity of the protein for nickel sepaharose was possibly due to inaccessibility of the tag due to inopportune protein folding. Affinity chromatography under denaturing conditions was attempted but did not improve yield, suggesting that the tag remained inaccessible even after the protein had been denatured.

Cloning of the CD0386 into pTAC-MAT1 to allow expression of CD0386 with a N-terminal metal affinity tag significantly increased yield from metal affinity purification and when this was combined with a subsequent ion exchange chromatography step, a purity and yield was obtained which was considered satisfactory for the purposes of immunisation for raising of antiserum. However, significant contamination with higher molecular weight aggregates was observed, which could not be removed by size exclusion chromatography.

4.4.4 Purification of CD0386N

Due to the large size of CD0386 and the considerable difficulties in its purification, a smaller fragment was deemed necessary for crystallisation experiments. To this end, a gene fragment encoding a 61kDa N-terminal fragment of CD0386 was cloned into pET28a fused with an N-terminal thrombin-cleavable hexahistidine tag. Purification of CD0386N by metal affinity chromatography resulted in much higher purity than was obtained with full-length CD0386. Also, unlike the full length protein, it was possible to remove the high molecular weight aggregates by size exclusion chromatography, resulting in a significantly more homogenous product.

SDS-PAGE of CD0386N results in three bands, one at the expected mass of 61kDa and two bands at ~40kDa and ~20kDa. However, mass spectroscopy does not reveal major contaminants at these masses and the purified protein was obtained from a single, well-shaped peak, suggesting a monodisperse species. Given that the summed masses of the unexpected bands equal that of CD0386, it possible that these two bands correspond to breakdown products, and based on the above evidence that these contaminants are absent in the raw purified product, it is hypothesised that the protein is susceptible to breakdown during preparation for SDS-PAGE. This could be confirmed by excision of the unknown bands and mass spectroscopy of the excised protein. For the purpose of this purification it is satisfactory that both mass spectroscopy and size exclusion chromatography indicate a predominantly monodisperse species.

Chapter 5 – Characterisation of a *C. difficile* Sortase Knockout

5.1 Introduction and Aims

5.1.1 Introduction to Genetic Manipulation of *C. difficile*

For many years, *C. difficile* was considered refractory to genetic manipulation. Early efforts focussed on transposon mutagenesis, with the demonstration that DNA within transposon Tn916 could be inserted into the genome following conjugative transfer of a Tn916-bearing plasmid from *B. subtilis* (Mullany *et al.*, 1994). Initial experiments in *C. difficile* CD37 indicated that insertion of Tn916 occurs at defined hotspots (Mullany *et al.*, 1991) which would make it of limited use for mutagenesis, and although later experiments in *C. difficile* 630 and R20291 resulted in a random and multiple insertion pattern (Hussain, 2005; Mullany *et al.*, 2012) it remains of limited use for mutagenesis experiments.

A major obstacle to the development of genetic manipulation in *C. difficile* was the lack of a suitable vector system. In *C. difficile*, as in many species, introduction of exogenous plasmid vectors is hindered by the presence a restriction/methylation system which protects the host by selectively cleaving exogenous DNA. By characterising the restriction/methylation system of *C. difficile* and engineering a plasmid without recognition sites for its restriction enzyme CdiI (Purdy *et al.*, 2002), it has become possible to efficiently transfer shuttle vectors from *E. coli* to several strains of *C. difficile*.

With plasmid vectors now readily available, the most common genetic manipulation technique currently utilised in *C. difficile* is insertional mutagenesis with mobile bacterial introns. Self-splicing mobile introns are present in many bacterial, plant and mitochondrial genomes, and of particular interest are the mobile group II introns. These encode an intron encoded protein (IEP) and a ribozymic RNA component. The IEP binds to the RNA to form a ribonucleoprotein complex (RNP), activating ribozymatic activity by which the RNA splices itself into the sense strand of the target DNA (Cousineau *et al.*, 1998). The IEP then cleaves the antisense strand and reverse transcribes the integrated RNA to generate a single stranded DNA (Zimmerly *et al.*, 1995). The process by which the sense strand is synthesed to repair the site is dependent on host DNA polymerases rather than any activity of the IEP (Smith *et al.*, 2005).

The overall process activity is known as 'retrohoming' (See Figure 5.1.1), and because target sequence recongition is co-operative between the protein and RNA components, the sequence of

the RNA component can be engineered to promote integration of the intron at specific sites (Mohr *et al.*, 2000). In addition, genetic elements can be introduced into the intron without affecting mobility (Smith *et al.*, 2005), allowing selectable markers to be employed. Of particular utility are Retrotransposition Activated Markers (RAMs), consisting of an antibiotic resistance gene interrupted by a group I intron. Splicing out of the group I intron occurs only when the RAM is integrated into chromosomal DNA, so acquisition of antibiotic resistance can be used to select group II intron integrants.

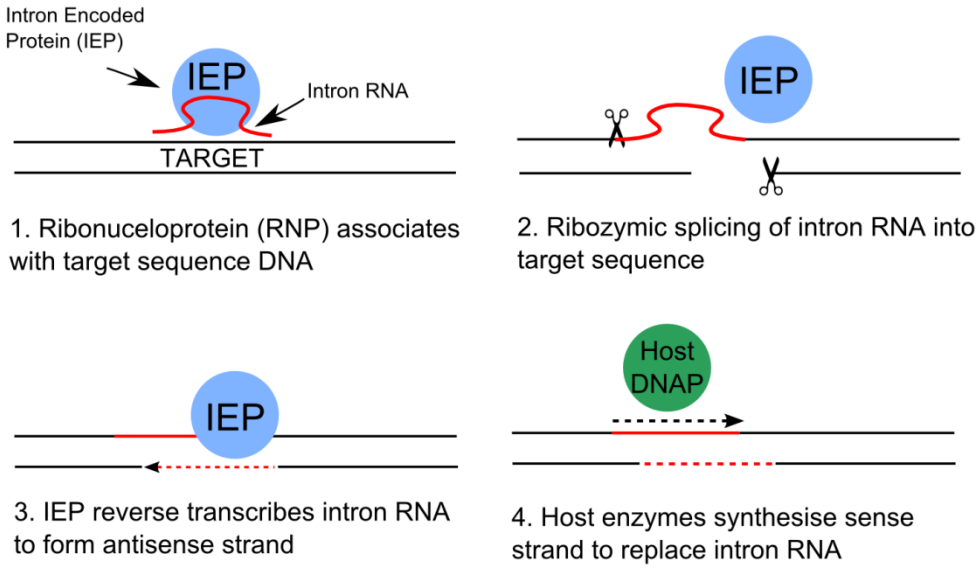


Figure 5.1.1 – Schematic of Retrohoming by a Group II Intron. Following recognition of the target DNA by the intron ribonucleoprotein, the intron RNA splices itself into the target. The Intron Encoded Protein reverse transcribes the intron sequence into the opposite strand, following which host enzymes replace the intron RNA with DNA.

The use of retargeted *Lactococcus lactis* Ll.LtrB intron for targeted mutagenesis became widespread after commercialisation by Sigma Aldrich under the name of 'Targetron', and an adaptation of the system for use in *C. difficile* was published in 2007 (Heap *et al.*, 2007). The system employs the Ll.LtrB intron within a conjugative shuttle vector pMTL007 which is capable of independent replication in either *E. coli* or *C. difficile* and is competent for conjugation between the two. Thus retargeting of the plasmid-borne intron can be performed in *E. coli* before transfer into *C. difficile* by conjugation. The intron encoded by pMTL007 includes an erythromycin resistance RAM, so an erythromycin sensitive strain must be used. In the case of *C. difficile*, an erythromycin sensitive mutant of strain 630 (*C. difficile* 630Δ*erm*) is typically employed.

5.1.3 Aims

The aims of the work described in this chapter were to investigate the *C. difficile* sortase by the following means:

1. Analysis of the *C. difficile* genome to identify predicted proteins with characteristics of sortase-anchored proteins and to provide a putative summary of their functions.

2. Construction of a a *C. difficile* sortase (*srtB*) gene mutant by intron mutagenesis and confirmation of a loss of SrtB from the cell surface.

3. Characterisation of the sortase mutant with particular regard to the localisation of the putative wall-anchored proteins identified by *in silico* analysis.

5.2 Methods

5.2.1 *In Silico* Analysis of The Putative *C. difficile* Class B Sortase

For placement of the putative *C. difficile srtB* gene (gene locus CD2718) within a phylogenetic tree of sortases, five representative sortase genes were chosen from each class and a distance tree drawn using a maximum likelihood method using the online pipeline provided by the Centre National de la Recherche Scientifique, France (www.phylogeny.fr, Dereeper et al. 2008). For identification of conserved amino acids within the sequence of SrtB, amino acid sequence alignments were performed using an implementation of ClustalW algorithm hosted by the European Bioinformatics Institute (www.ebi.ac.uk/Tools/msa/clustalw2/, Larkin et al. 2007).

5.2.2 Identification and Analysis of Putative Sortase Substrates

The genome of *C. difficile* 630 was searched for proteins containing an [SP]PxTG motif using the Basic Local Alignment Search Tool (BLAST) of the National Centre for Biotechnology Information (http://blast.ncbi.nlm.nih.gov/, Altschul et al. 1990). The resulting list of [SP]PxTG-containing proteins was further refined by removal of proteins which did not contain N-terminal secretion signal as predicted by the SignalP 4.0 program (www.cbs.dtu.dk/services/SignalP, Petersen et al. 2011), or those in which the [SP]PxTG motif was not within 50 amino acids of the C- terminus. To ascribe putative functions to the products of the identified genes, conserved domains were identified using the NCBI Conserved Domain Search (Marchler-Bauer and Bryant, 2004) and secondary structure prediction and secondary structure homology searches were performed using the Phyre2 server (www.sbg.bio.ic.ac.uk/~phyre/, Kelley & Sternberg 2009)

5.2.3 Generation of Targeted Insertion Knockout Mutants

Targets for insertion within the *srtB* gene (gene locus CD2718) were chosen using the online tool provided by Sigma for the targetron system (http://www.sigma-genosys.com/targetron/). The targets sites identified are shown in Table 5.2.1.

Insertion site	Sequence 5' exon	Sequence 3' Exon	Score
354/255a	CTTTCTGTTGAGAATACAAATATAAATTAT	CCAGTTGTACAATCT	8.3
510|511s	GTAGAATCTGATTATGATTATCTAAAAACA	AATTTTAATAATGAA	8.19
514|515a	TGTAATTTTGATAATCAGATTCATTATTAA	AATTTGTTTTTAGAT	8.07
206|207a	TTATCTTTAGATTGTACAACTGGATAATTT	ATATTTGTATTCTCA	7.13
213|214s	ATTTAGGTTGTTGAACATAGTTTTATTTTT	CATATTATGTCCATA	6.32

Table 5.2.1 Hypothetical Ll.LtrB intron insertion sites within CD2718 (*srtB*) as predicted by the Sigma Targetron Tool. Insertion site loci are expressed as base pairs from the start of the open reading frame. Sense or antisense insertion is indicated by 's' or 'a' respectively after the site locus.

pMTL007 plasmids were retargeted for the sites listed in Table 5.2.1 and synthesised by DNA2.0 on supply of the sequence for the retargeting region, generated by replacing the retargeting sequences of the intron with the sequences as provided by the Sigma targetron tool. *E. coli* CA343 was transformed with retargeted pMTL007 plasmids by electroporation.

A culture of *E. coli* CA343 pMTL007 harbouring the retargeted intron was grown overnight in LB to an optical density of 0.6 at 600nm. Cells were harvested from 1ml of broth by centrifugation at 3,000g for 30mins and resuspended in 100µl of a culture of *C. difficile* 630Δ*erm* grown to an optical density of 0.6 at 600nm, resulting in a 100:1 ratio of *E. coli* to *C. difficile*. The culture mixture was spread on a 22µm nitrocellulose filter paper (Sartorius AG) placed in the centre of an FAA plate. After 3 days of incubation, filter papers were removed, washed with sterile pre-reduced phosphate buffered saline and transconjugants selected by spreading of the resulting suspension spread on *C. difficile* selective agar plates (E&O Laboratories, Bonnybridge UK) supplemented with 15µg/ml thiamphenicol.

Integrants were selected by restreaking transconjugant colonies onto BHI agar supplemented with 5µg/ml erythromycin. Colonies were picked and colony PCR performed with primers flanking the group I intron (Table 3.2.3 ErmR and ErmF) to confirm that the erythromycin resistant phenotype was due to splicing out of the group I intron within the ErmRAM. To confirm specific integration

of the intron into the intended site, PCR was performed with primers flanking the gene of interest (Table 3.2.3, CD2718FlankF and CD2718FlankR) and the products sequenced.

5.2.4 Determination of Growth Rate and Toxin Production of *C. difficile* 630 *srtB::erm*

Intracellular and extracellular toxin titres were determined by an Enzyme Linked Immunosorbent Assay (ELISA) employing sheep antitoxin polyclonal antibodies as a capture step. Growth rate data were collected simultaneously with toxin sampling in order to ensure equivalence of growth between strains.

Two 14ml aliquots of reduced BHI were inoculated with either *C. difficile* 630 or *C. difficile srtB::erm* and grown overnight. Three 1ml aliquots were taken from each, centrifuged at 4,000g for 15mins and the pellets resuspended in 1ml fresh BHI, a step neccesary to remove any toxin produced in the overnight culture. The resuspended culture was then used to inoculate 25ml of fresh BHI. At time points 0h, 8h, 24h, 30h, and 48h, 2ml samples were taken from each culture. Following measurement of optical density at 600nm, culture samples were centrifuged at 10,000g for 10mins and the supernatant decanted and frozen at -80°C. Pellets were resuspended in 1ml BHI, sonicated on ice for 20s then stored at -80°C until analysis by ELISA.

Sheep polyclonal anti-TcdA or anti-TcdB antibodies were diluted to a concentration of 5µg/ml and a volume of 100µl applied to a high binding 96 well plates (Corning). Plates were incubated at 37°C for 90mins, then washed thrice in PBS-T (1x PBS, 0.1% Tween). Plates were blocked by incubating with with 150µl blocking buffer (PBS-T + 5% foetal calf serum) for 1hr at 37°C and then washed thrice with PBS-T. Dilution series of standards and samples were assembled in non-binding 96-well plates (Corning) Standard preparations of TcdA or TcdB were diluted to 1µg/ml in blocking buffer and 300µl applied to wells 1A and 1B of the plate. Samples for analysis were applied in triplicate to the remaining wells of column 1 in a volume of 300µl. The remainder of the wells on the plate were filled with 150µl of blocking buffer. A multichannel pipette was used to perform a 1 in 2 serial dilution series across the plate. From this dilution plate, 100µl of was transferred to the assay plate and incubated for 1hr at 37°C before washing thrice with PBS-T.

Biotinylated sheep anti-TcdA or anti-TcdB polyclonal antibodies were diluted in blocking buffer to either 1µg/ml for TcdA or 10µg/ml for the TcdB. 100µl added to all wells of the appropriate plate and incubated for 1hr at 37°C before washing thrice with PBS-T. A streptavidin horseradish peroxidase conjuagate was diluted 1/1000 in blocking buffer and 100µl added to all wells and incubated for 10mins, followed by washing 3x with PBS-T. 100µl of 3,3',5,5'-Tetramethylbenzidinem (TMB) substrate was added to each well, incubated at 37°C for 10mins.

100µl of TMB stop reagent was added to each well and absorbence at 450nm measured with a plate spectrophotometer (Tecan), with reference readings taken at 620nm. All washing steps were performed using a 12-head automated microplate washer (Thermo)

5.2.5 Immunofluorescence with anti-CD0386 antibodies

Unless otherwise noted, centrifugation steps described below were for 10mins at 5,000g and where washing consisted of is described, this consisted of three resuspensions in 1ml PBS followed by three centrifugation steps at 5000g..

C. difficile was grown to an optical density of 1.0 at 600nm. Aliquots of 1.5ml culture were centrifuged and washed. The pellet was resuspended in 100µl sterile water, added to 900µl ice cold methanol, incubated for 10mins at room temperature with rotational incubation and then washed thrice. Cells were resuspended in PBS supplemented with 1mg/ml Bovine Serum Albumin (BSA), either with or without 10µg/ml of purified primary antibody, and incubated with rotation for 1hr at room temperature. After washing with PBS, cells were resuspended in 1ml PBS 1mg/ml BSA, with 10µg/ml anti-rabbit Alexa 488 conjugated antibody (Thermo), incubated with rotation for 1hr. After washing in PBS, cells were resuspended in distilled water. The suspension was spotted onto a glass slide and air dried in darkness before inspection with an MWG EVOS fluorescence microscope using either visible light or UV light with the manufacturer's GFP filter (transmission peak 525/550nm).

5.2.6 Automated Counting of Fluorescent Cells

Samples of *C. difficile* 630 and 630 *srtB::erm* were prepared as in Section 4.2.5. To ensure blind counting of fluorescent cells, the CellC image analysis program was utilised (Selinummi *et al.*, 2005). Fields were selected blind from three biological replicates and images were taken in duplicate visible/UV. The software was configured to run in batch mode and to count cells as fluorescent when appearing in both the visible and UV images. The mean number of fluorescent cells was divided by the mean number of total cells to obtain the proportion of cells that were fluorescent.

5.2.7 Fractionation of *C. difficile*

Cells were fractionated using a method adapted from previous reports of proteomic analysis of *C. difficile* (Wright *et al.*, 2005). 1ml aliquots of *C. difficile* 630 and 630 *srtB::erm* were retrieved from storage at -80°C and used to inoculate 14ml of BHI and incubated anaerobically overnight at 37°C. Four sets of 3% subcultures were made (ie. four cultures of each knockout) in 50ml of fresh BHI and incubated anaerobically at 37°C At 5hrs, 24hrs and 48hrs, the following fractionation protocol was performed on both the wild type and knockout.

Culture aliquots of 1ml were pelleted by centrifugation at 3,000g for 10mins, resuspended in 20ml Tris Sucrose Buffer (TS – 100mM Tris pH6.9, 10mM $MgCl_2$, 0.5M Sucrose) and centrifuged once again at 3,000g for 10mins. Pellets were resuspended in 2.4ml of digestion buffer (DB – TS Buffer + 1mg/ml lysozyme and 60µg/ml mutanolysin) and incubated at 37°C for 60mins with rotational agitation. Protoplasts were recovered by centrifugation at 10,000g for 10mins. The supernantant, composed of the wall fraction, was decanted and frozen at -80ºC.

The protoplast pellet was resuspended in 1ml protoplast lysis buffer (10mM Tris pH7.4, 100mM NaCl, 10mM $MgCl_2$), incubated for 60mins on ice and freeze-thawed twice at in a dry ice/ethanol bath. Membrane fraction was recovered by centrifugation at 10,000g for 10mins. The supernatant, composed of the cytoplasmic fraction, was decantedand frozen at -80ºC . The pellet was resuspended in 1ml of TS buffer and sonicated 3 x 10s until homogenous. To ensure equivalence between samples, protein concentration was determined by BCA assay. Analysis by SDS PAGE and western blotting was performed as detailed in Section 3.3.

5.2.8 Immunoprecipitation of CD0386 from *C. difficile* wall fraction

Immunoglobulin G (IgG) was purified from anti-CD0386 rabbit serum using a Protein A sepharose column (GE healthcare) according to the manufacturers instructions. 20µg of IgG was coupled to a direct immunoprecipitation column as supplied in Pierce Direct Immunoprecipitation kit, following which the column was incubated at 4ºC overnight with either 600µl of wall fraction, 60µl of wall fraction diluted to a volume of 600µl with the Pierce immunoprecipitation buffer or 20µg of recombinant CD0386 in a volume of 600µL buffer.

5.2.9 Preparation of purified *C. difficile* peptidoglycan

Peptidoglycan purification was attempted by the 'boiling SDS' method, whereby culture was added dropwise to an aqueous solution of 8% SDS at 95ºC. After heating for 2hrs at 95ºC, insoluble material was recovered by centrifugation at 10,000g and resuspended in a volume of water equal that that of the original culture volume before being added dropwise to a fresh solution of 8% SDS. This washing process was repeated thrice. Fractions were collected throughout and analysed by SDS-PAGE.

5.2.10 Challenge of Hamsters with *C. difficile* 630 *srtB::erm*

The animal procedures described in this study were perfomed by staff of Public Health England, Porton Down. Procedures were approved by a local ethical committee and carried out under a UK Home Office-approved Project Licence and in full compliance with the UK Animals (Scientific Procedures) Act 1986.

To prepare spores for challenge experiments, 12 Facultative Anaerobe Agar (FAA) plates were inoculated with overnight growth of either *C. difficile* 630 or *C. difficile* 630 *srtB::erm* in 5ml sBHI. After 14 days of anaerobic incubation at 37°C, colonial growth was scraped with a sterile swab and resuspended in 15ml Dulbecco's Modified Eagle's Medium (DMEM). The resuspension was heat-shocked at 62°C for 40mins to kill vegetative cells, aliquoted and stored at -80°C. Spore counts were determined by plating of tenfold serial dilutions on FAA agar.

Female syrian hamsters (80-100g) were housed in pairs in isolator cages with filter lids to minimise contamination between groups. Clindamycin was administered to all groups at a dose of 2mg in 0.2ml sterile H_2O two days prior to challenge. Group 1 (n=10) was challenged orally with 10^3 *C. difficile* 630 spores, Group 2 (n=10) was challenged with 10^3 *C. difficile* 630 *srtB::erm* spores while Group 3 (n=6) remained unchallenged.

All animals were weighed daily and monitored six times per day throughout the experiment. Hamsters were scored (0-3) on diarrhoea, weight loss, lethargy, and tender abdomen. Hamsters exhibiting advanced symptoms of disease were euthanized humanely.

5.3 Results

5.3.1 *In Silico* Analysis of The Putative *C. difficile* Sortase

A maximum likelihood phylogenetic tree (Figure 5.3.1) places the gene locus CD2718 (*srtB*), within a branch containing the class B enzymes.

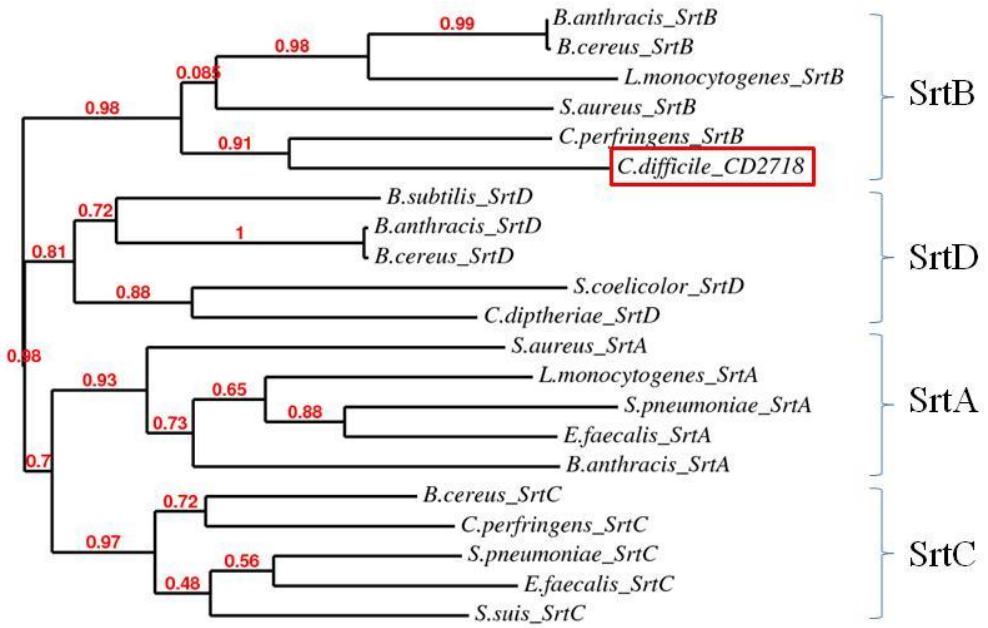


Figure 5.3.1 – Phylogeny of *C. difficile* CD2178 (*srtB*). Phylogram showing placement of the *C. difficile* CD2718 (*srtB*) gene (boxed) within a tree of selected sortases from the four sortase classes.

Multiple sequence alignment (Figure 5.3.2) reveals several motifs which are conserved in the class B sortases, including a TLxTC motif which is characteristic of sortase enzymes containing the active site cysteine residue equivalent to Cys184 in *S. aureus* SrtA, a conserved arginine downstream of the TLxTC motif in a similar position to Arg233 of *S. aureus* SrtA and a 'YGH' motif containing the structural equivalent of the *S. aureus* SrtA His120 (Zhang *et al.*, 2004)

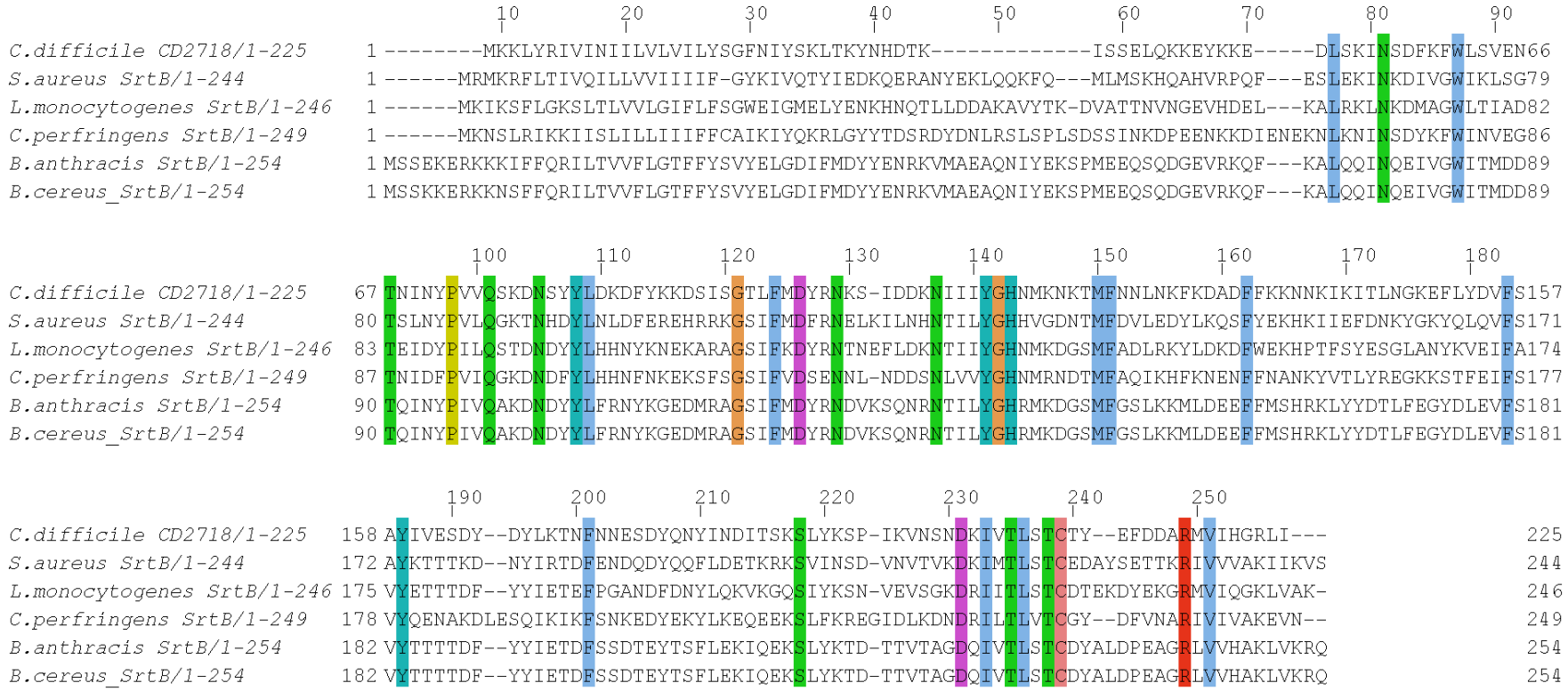


Figure 5.3.2 *In Silico* Analysis of the Putative *C. difficile* Sortase. A) Multiple alignment of selected SrtB protein sequences and the putative *C. difficile* SrtB by the 'MUSCLE' method. Strictly conserved residues are highlighted, gaps are indicated by a dash (conserved residues coloured by properties).

5.3.2 *In Silico* Identification of the Putative *C. difficile* Sortase Substrates

A BLAST search to identify hypothetical *C. difficile* 630 proteins containing an [SP]PxTG motif led to a preliminary list of 25 proteins. Of these, 15 were eliminated because they did not contain an N-terminal secretion signal peptide as determined by the SignalP program. The remaining proteins were examined manually, and a further 3 eliminated either because the [SP]PxTG motif was too far (<50aa) from the C terminus or because the C-terminus was not rich in positively charged amino acids. The remaining seven proteins, containing all of the characteristics of sortase substrates, are listed in Table 5.3.1 accompanied by gene annotations and results of conserved domain searches while their C-terminal cell wall sorting signals are shown in Figure 5.3.3.

ORF Name	Mass (kDa)	Gene Annotation	Predicted Conserved Domains (CDs)	E-Value for CDs
CD0183	37.0	Cell Wall Hydrolase	Pfam 877 (Unknown Function) Pfam08239 (Bacterial SH3 domain) COG0791 (Cell Wall Hydrolase)	5.66e-36 1.12e-05 1.94e-37
CD0386	111.6	Putative collagen-binding surface protein	COG4932 (Predicted Membrane Protein)	1.14e-10
CD2537	68.2	Membrane-associated 5'-nucleotidase	CD07407 (N-terminal metallophosphatase domain) Pfam02872 (5' nucelotidase, C-terminus) COG0737 (5'nucleotidase/2',3'-cyclic phosphodiesterase)	1.97e-101 8.37e-59 5.92e-94
CD2768	24.9	Cell Wall Hydrolase	Pfam 877 (Unknown Function) Pfam08239 (Bacterial SH3 domain) COG0791 (Cell Wall Hydrolase)	2.42e-35 7.2e-11 5.11e-35
CD2831	107.7	Putative Adhesin	Pfam05738 (Cna B Domain) CL05349 (Collagen Binding Domain)	2.33e-06 3.79e-03
CD3246	79.9	Surface Protein	Pfam06346 (Phormin Homology Region)	5.39e-05
CD3392	111.6	Cell Surface Protein	Pfam05738 (Cna B Domain)	1.47e-03

Table 5.3.1 – The Putative *C. difficile* sortase substrates. Identification of *C. difficile* proteins fulfilling the criteria for sortase substrates, containing a C-terminal [SP]PxTG motif and a N-terminal secretion signal.

```
CD0183 HSPSTGKTVSVTSINSSYYTARFVTAKRIL
CD0386 DSPKTGDNTNLYGLLALLLTSGAGLAGIFFYKRRKMKKS
CD2537 KSPKTGDLGFSNSIIIFIVSSTLICLLNFNQKELKDKKSK
CD2768 HSPQTGDVVKVTSMAPGTNYARRLITATRVLQ
CD2831 VPPKTGDSTTIIGEILLVIGAIVGLIVLRRNKNTN
CD3246 KSPKTGDETQLMSYVFISVIAICGLAYQCKIKRN
CD3392 DSPKTGDSTNLMAFIVMLLVSGGGLAGTYLYKRRKMKKS
```

Fig 5.3.3. Alignment of the C-termini of the Seven Putative *C. difficile* Sortase Subtrates. The C-termini of the proteins listed in Table 5.3.1 are shown. The [SP]PxTG motif has been underlined and positively charged residues are highlighted in red.

5.3.3 Phyre2 Analysis of Putative *C. difficile* Sortase Substrates

To aid in ascribing putative functions to the putative sortase substrates, their sequences were submitted to the Phyre2 server, which performs secondary structure prediction followed by searching of PDB entries to predict structural homologues. With the exception of CD3246, for which Phyre2 predicted very low secondary structure, alignment hits with >98% confidence were obtained for all proteins. The results of the Phyre2 search are tabulated in Table 5.3.2.

CD0386:

Template	Description	Species	%ID	Coverage
3kptA	BcpA Pilin	*B. cereus*	22	37%
2ww8A	RgrA Pilin	*S. pneumoniae*	21	50%
2xicB	Pilus Adhesin	*S. pyogenes*	22	33%

CD2768:

Template	Description	Species	%ID	Coverage
3npfB	Dipeptidyl peptidase vi	*Bacteroides ovatus*	35	80%
2fg0B	γ-d-glytamyl-diamino acid endopeptidase	*Nostoc punctiforme*	27	80%
3h41A	Nlp/p60 family protein	*B. cereus*	25	82%

CD2537:

Template	Description	Species	%ID	Coverage
2wdfA	Sulfur oxidation SoxB	*Thermus thermophilus*	23	79%
2z1aA	5’ nucelotidase	*Thermus thermophilus*	30	78%
3ivdA	Put. 5’ nucelotidase	*E. coli*	28	78%

CD3392:

Template	Description	Species	%ID	Coverage
3kptA	BcpA Pilin	*B. cereus*	21	36%
2ww8A	RgrA Pilin	*S. pneumoniae*	23	50%
2xicB	Pilus Adhesin	*S. pyogenes*	20	36%

CD0183:

Template	Description	Species	%ID	Coverage
3h41A	Nlp/p60 family protein	*B. cereus*	30	85%
3npfB	Put. Dipeptidyl peptidase vi	*B. ovatus*	26	82%
2fg0B	γ-d-glytamyl-diamino acid endopeptidase	*N. punctiforme*	26	65%

CD2831:

Template	Description	Species	%ID	Coverage
3irpX	UafA adhesin functional region	*Staphylococcus saprophyticus*	16	63%
2xtlB	Major pilin	*Streptococcus agalacticae*	14	49%
3kptA	BcpA Major pilin	*Bacillus cereus*	21	18%

Table 5.3.2 Phyre2 Analysis of Putative *C. difficile* Sortase Substrates. Details of the three highest ranking predicted secondary structure alignments are shown for each protein. CD3246 is omitted due to lack of significant predicted secondary structure.

5.3.3 Generation of *C. difficile* SrtB Gene Knockout

A total of five retargeted introns were introduced into *C. difficile* 630Δ*erm* by conjugation from *E. coli* CA343. In all cases, colonies were isolated on thiamphenicol supplemented plates, indicating that conjugation of the plasmid had been successful. For introns retargeted at sites 354, 510, 514, and 206, no colonies were isolated on erythromycin-supplemented plates, indicating that integration of the intron had not occurred. The entire protocol was repeated twice with each intron before unsuccessfully retargeted introns were abandoned.

At site 213, several colonies were isolated on erythromycin supplemented plates. PCR amplification of gDNA from these integrants was performed with primers flanking the *srtB* gene. This revealed a characteristic ~1.9kb insertion within the *srtB* gene (Figure 4.3.4A). The insertion was then fully characterised by sequencing of the intron junctions (Figure 5.3.5). To confirm lack of SrtB protein in the knockout strain, western blots of cell fractions was performed as in Section 4.2.7 and probed with rabbit serum raised against recombinant SrtB (Figure 5.3.4B). The knockout strain is henceforth described as *C. difficile* 630 *srtB::erm*.

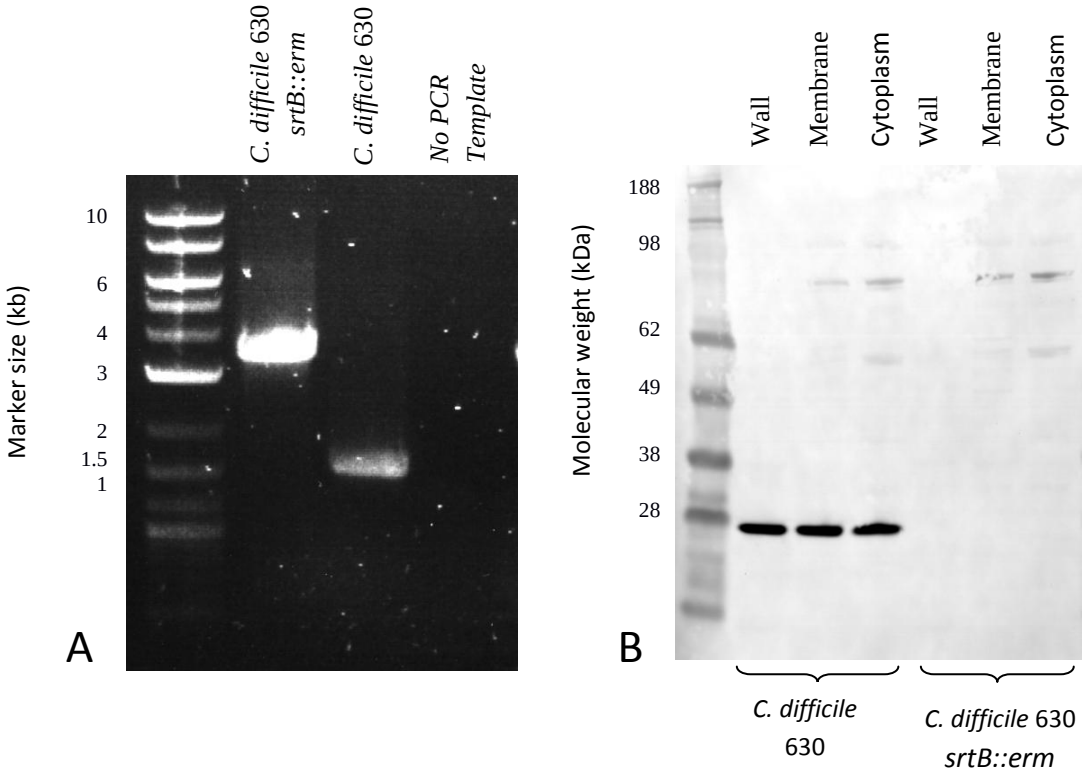


Fig 5.3.4 Confirmation of Interruption of *srtB* by Intron Mutagenesis. A) DNA fragments generated by PCR amplification utilising primers flanking the *srtB* gene, demonstrating an insertion of ~1.9kb in strain 630 *srtB::erm* corresponding to the size of the Ll.LtrB intron. B) Western blotting of cell fractions from 630 and 630 *srtB::erm* with antibodies raised against recombinant SrtB.

5' of CD2718 (*srtB*)

C. difficile 630

ATCATGATACTAAAATATCCTCAGAACTTCAAAAAAAAGAATATAAAAAAGAGGATTTATCAAAGATTAACAGTGATTTTAAATTTTGGCTTTCTGTTG
AGAATACAAATATAAATTATCCAGTTGTACAATCTAAAGATAATTCTTACTATCTTGACAAAGATTTTTATAAAAAGATTCTATTTCTGGTACACTGTT
TATGGATTATAGAAATAAATCTATTGATGACAAAATATCATATATATGGACATA

C. difficile 630 *srtB::erm*

ATCATGATACTAAAATATCCTCAGAACTTCAAAAAAAAGAATATAAAAAAGAGGATTTATCAAAGATTAACAGTGATTTTAAATTTTGGCTTTCTGTTG
AGAATACAAATATAAATTATGTGCGCCCAGATAGGGTGTTAAGTCAAGTAGTTTAAGGTACTACTCTGTAGATACACAGAAAACAGCCAACCTAACCGA
AAAGCGAAAGCTGATACGGGAACAGAGCACGGTTGGAAGCGATGAGTTACCTAA

3' of CD2718 (*srtB*)

C. difficile 630

ATATCCTCAGAACTTCAAAAAAAAGAATATAAAAAAGAGGATTTATCAAAGATTAACAGTGATTTTAAATTTTGGCTTTCTGTTGAGAATACAAATATA
AATTATCCAGTTGTACAATCTAAAGATAATTCTTACTATCTTGACAAAGATTTTTATAAAAAAGATTCTATTTCTGGTACACTGTTTATGGATTATAGA
AATAAATCTATTGATGACAAAAATATCATAATATATGGACATAATATGAAAAATAAAACTATGTTC

C. difficile 630 *srtB::erm*

ACAATAACAGAGCCGTATACTCCGAGAGGGGTACGTACGGTTCCCGAAGAGGGTGGTGCAAACCAGTCACAGTAATGTGAACAAGGCGGTACCTCCCTA
CTTCACCCAGTTGTACAATCTAAAGATAATTCTTACTATCTTGACAAAGATTTTTATAAAAAAGATTCTATTTCTGGTACACTGTTTATGGATTATAGA
AATAAATCTATTGATGACAAAAATATCATAATATATGGACATAATATGAAAAATAAAACTATGTTC

Figure 5.3.5 Sequencing of the Interrupted *srtB* gene of *C. difficile* 630 *srtB::erm*. The sequence of the wild type is shown for reference. Underlining indicates Ll.LtrB intron DNA.

5.3.4 Growth and Toxin Production of *C. difficile* 630 *srtB::erm*

Growth curves were plotted for *C. difficile* 630 and 630 *srtB::erm* (Figure 5.3.6). No significant differences were observed, with identical growth rates and stationary phase density.

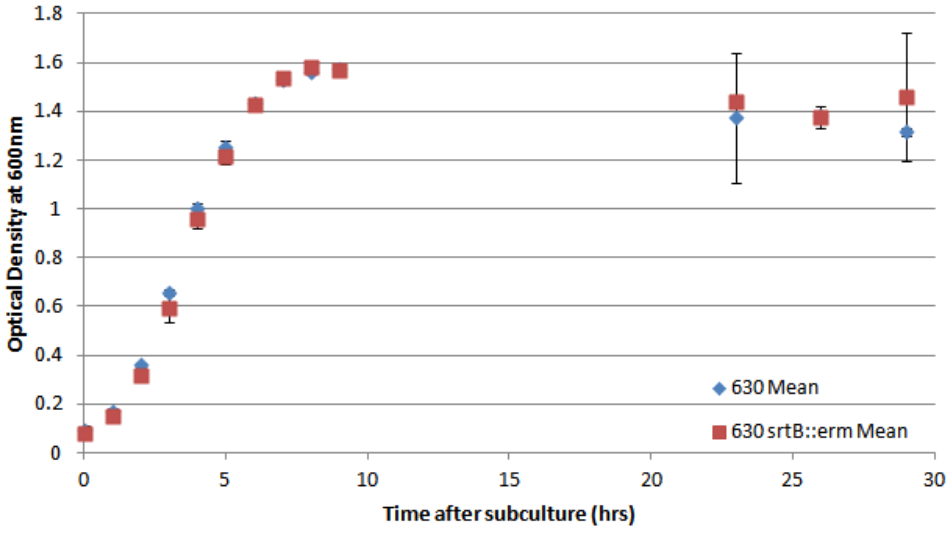


Figure 5.3.6 – Growth of *C. difficile* 630 and 630 *srtB::erm* over 30hrs. Triplicate culture samples were taken over 30hrs and absorbance at 600nm measured. Mean values (n=3) are plotted, error bars indicate one standard deviation of the mean.

Toxin titres, both intracellular and extracellular, were determined by a TcdA and TcdB capture ELISA as detailed in Section 4.2.4. Toxin A titres of 630 *srtB::erm* were statistically significantly higher ($p<0.05$) at 7, 24 and 50hrs intracellularly and at 24hrs extracellularly. Intracellular toxin B was higher in 630 *srtB::erm* at 7, 24 and 50hrs, of which 7hrs and 50hrs were statistically significant (Figure 5.3.7).

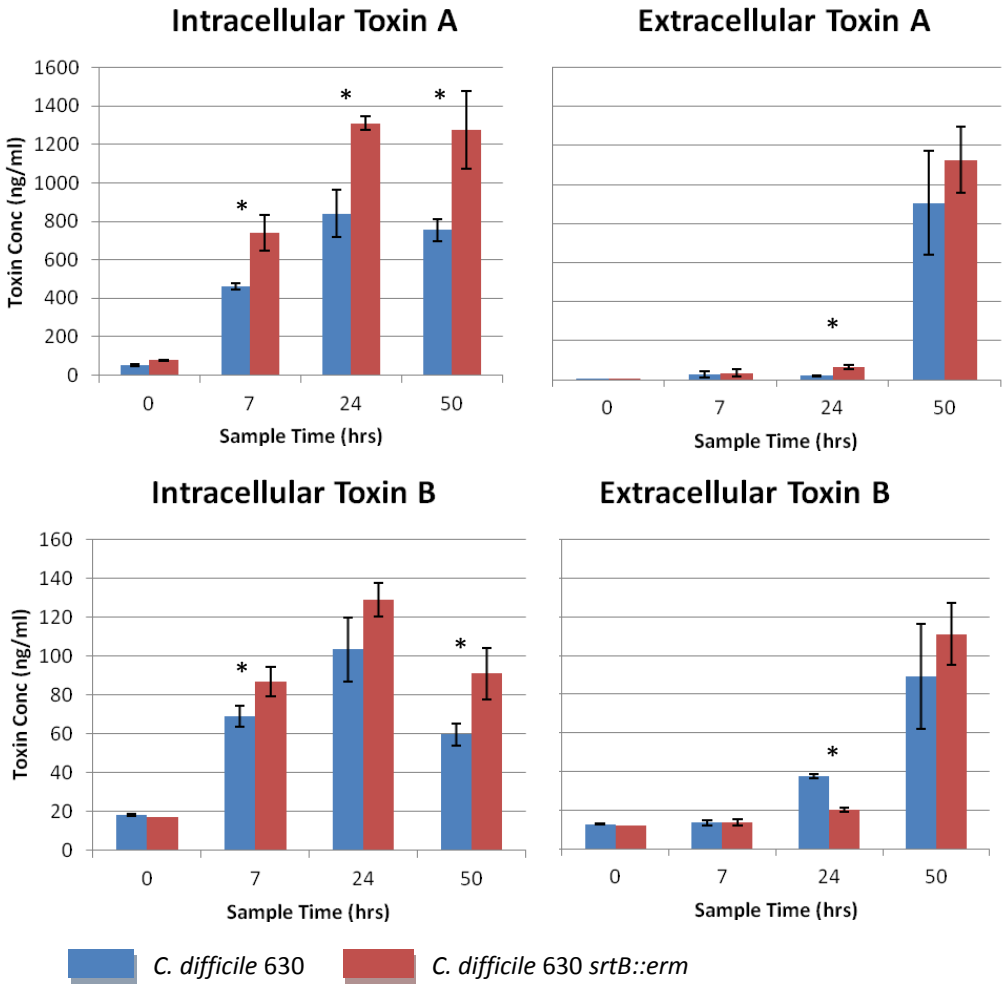
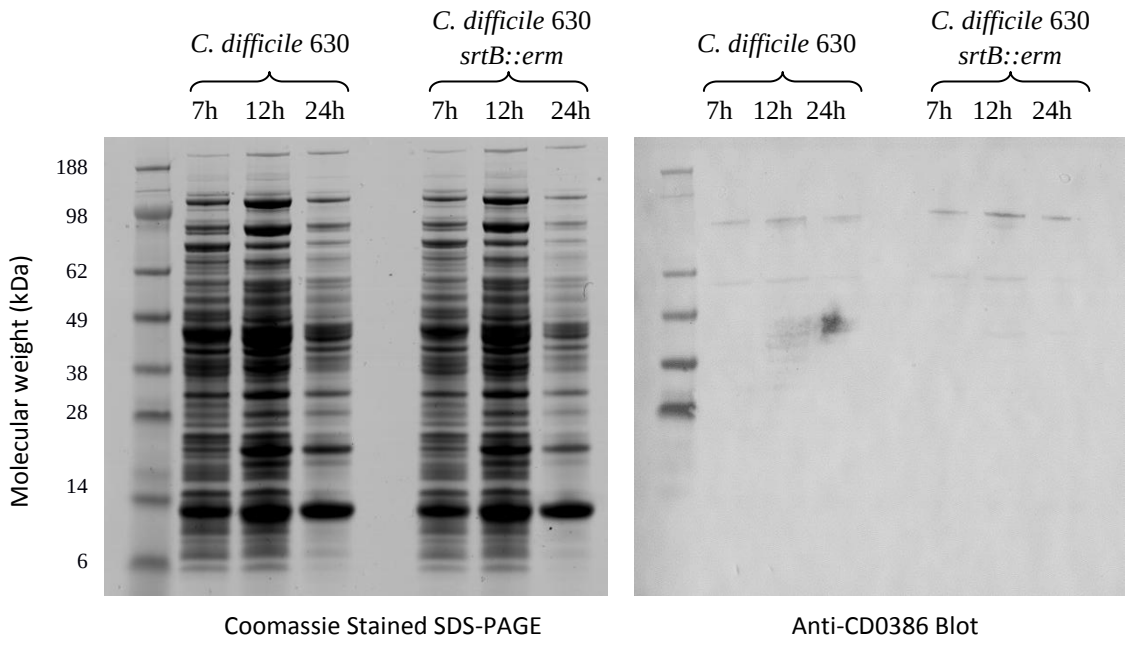


Figure 5.3.7. Intracellular and extracellular Toxin A and Toxin B titres of *C. difficile* 630 and *C. difficile srtB::erm*. Toxin Titres were determined by capture ELISA**.** Error bars display standard deviation (n=3). (*) indicates a difference in toxin titre statistically significant between strains ($p<0.05$) as determined by Students t-test.

5.3.5 Localisation of Sortase Substrate CD0386 in *C. difficile* 630 *srtB::erm*

By muramidase digestion of *C. difficile* cells and subsequent lysis and centrifugation of the resulting protoplasts, wall, membrane and cytoplasm fractions were obtained. Fractions were analysed by SDS-PAGE, blotted and probed with anti-CD0386 rabbit serum (Figures 5.3.8 and 5.3.9)

Cytoplasmic Fraction:



Membrane Fraction:

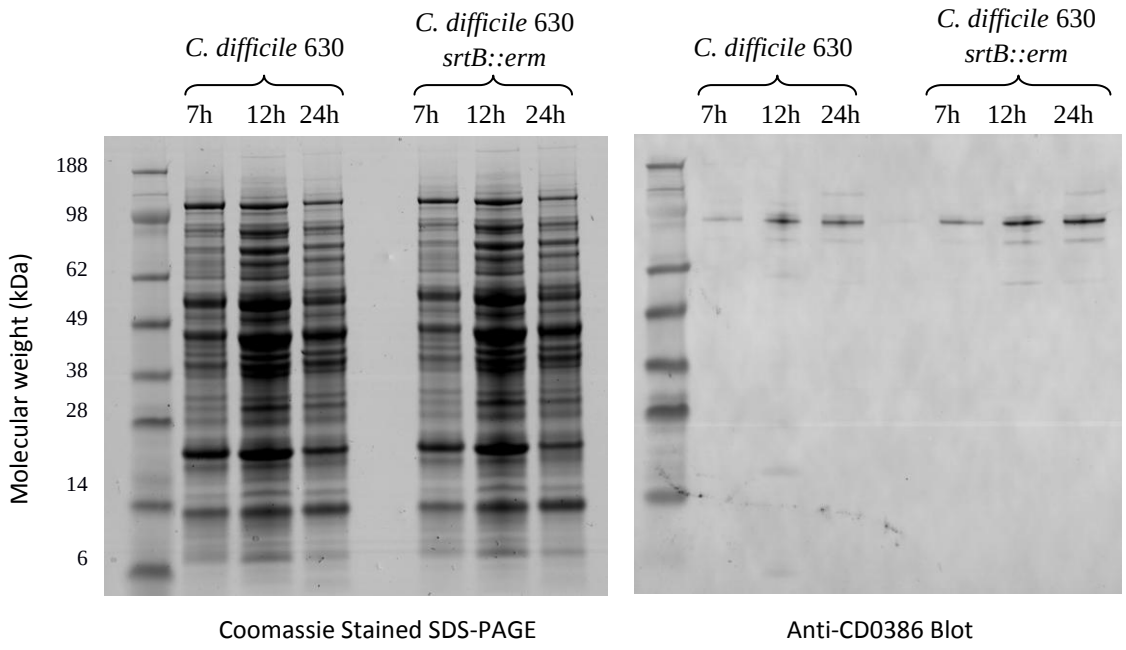


Figure 5.3.8 Cell Fractionation and Localisation of CD0386 (Cytoplasm and Membrane). Cytoplasmic and membrane fractions from *C. difficile* 630 and 630 *srtB::erm* after 7, 12 or 24hrs of growth were analysed by SDS-PAGE and either stained with coomassie or transferred to a nitrocellulose membrane and blotted with anti-CD0386 serum.

Wall Fraction:

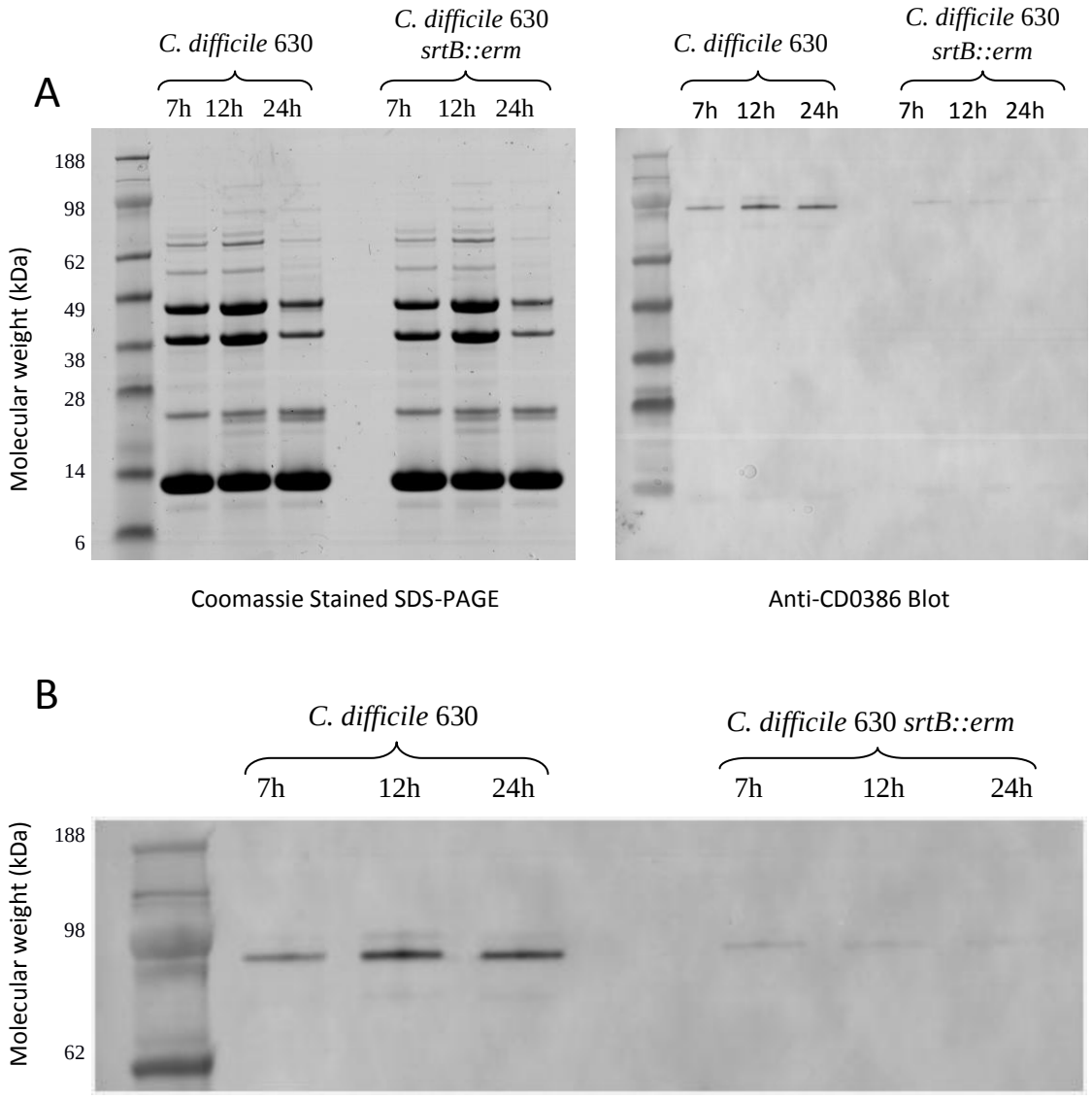


Figure 5.3.9 Cell Fractionation and Localisation of CD0386 (Wall). A) Wall fractions from *C. difficile* 630 and 630 *srtB::erm* after 7, 12 or 24hrs of growth were analysed by SDS-PAGE and either stained with Coomassie or transferred to a nitrocellulose membrane and blotted with anti-CD0386 serum B) Enlarged detail of the anti-CD0386 blotted membrane.

An unambiguous immunoreactive band corresponding to the approximate molecular weight of CD0386 is observed in the wildtype wall fraction, but is absent in the sortase knockout wall fraction. A fainter band of a slightly higher molecular weight is present in both the wild type and the sortase knockout. Anti-CD0386 reactive bands were observed in membrane and cytoplasmic fractions of both strains.

5.3.6 Immunofluorescent Staining of *C. difficile* with anti-CD0386 serum

Methanol-fixed *C. difficile* cells were incubated with anti-CD0386 antibodies and a fluorescent secondary antibody according to Section 3.2.5. Upon visualisation by UV microscopy, a small population of fluorescent cells was observed (Figure 5.3.10). Where no primary antibody was added, no fluorescence was observed, indicating that fluorescence was due to specific interactions of the CD0386 antibodies. The proportion of fluorescent cells in both *C. difficile* 630 and 630 *srtB::erm* samples was quantified by automated counting using the CellC software as described in Section 4.2.6. Significance testing using Students t test revealed no statistically significant difference between the two strains (Table 5.3.3)

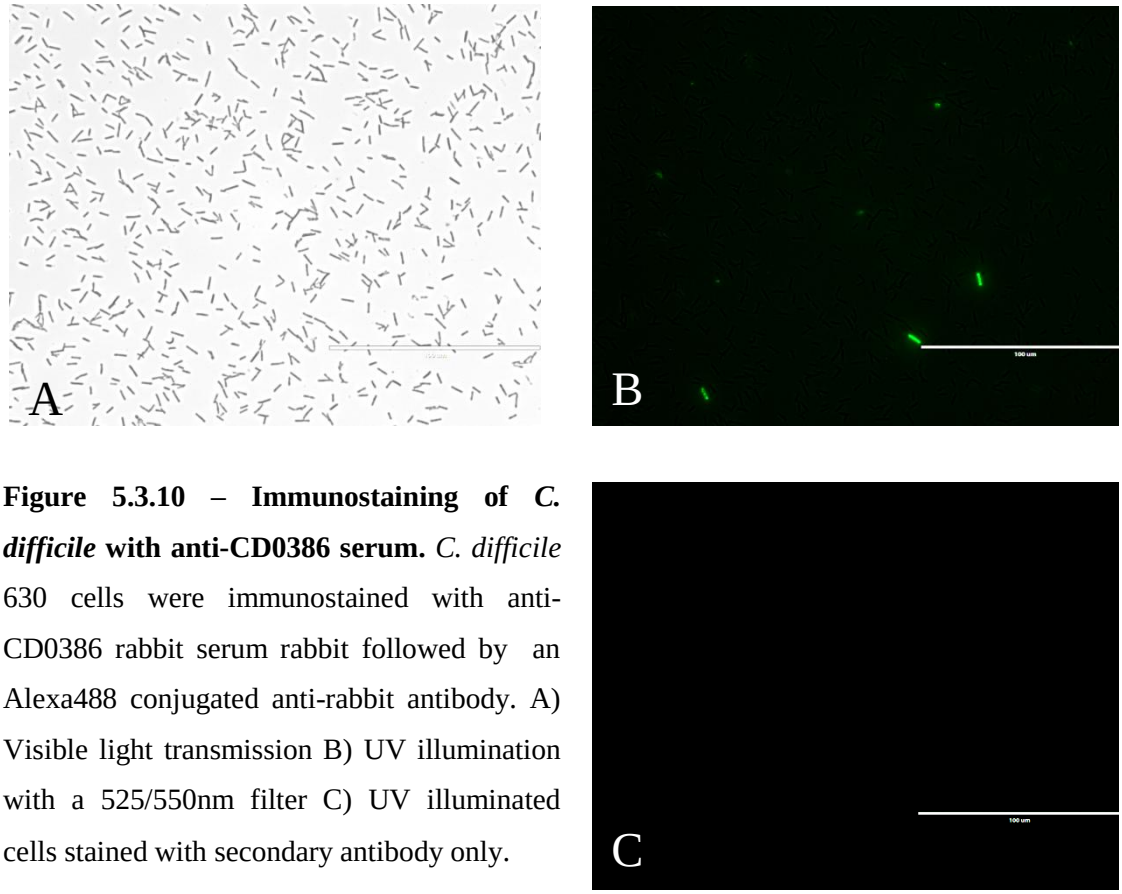


Figure 5.3.10 – Immunostaining of *C. difficile* with anti-CD0386 serum. *C. difficile* 630 cells were immunostained with anti-CD0386 rabbit serum rabbit followed by an Alexa488 conjugated anti-rabbit antibody. A) Visible light transmission B) UV illumination with a 525/550nm filter C) UV illuminated cells stained with secondary antibody only.

C.difficile 630:

	Total Cells	Fluorescent Cells	% of Cells Fluorescent
Set 1	1381	36	2.61%
Set 2	1931	30	1.55%
Set 3	1759	45	2.56%
Mean	1690	37	2.24%

C.difficile 630 *srtB::erm:*

	Total Cells	Fluorescent Cells	% of Cells Fluorescent
Set 1	3017	51	1.69%
Set 2	2249	40	1.78%
Set 3	1958	52	2.66%
Mean	2408	48	2.04%

Student's t Test:

	p-value (4 df)
Set 1	0.87
Set 2	0.40
Set 3	0.44

Table 5.3.3 – Proportion of fluorescent cells after anti-CD0386 Immunostaining. Students t test was used to test the hypothesis that there was no difference between the proportion of fluorescent cells of *C. difficile* 630 and *C. difficile* 630 *srtB::erm.*

5.3.7 Immunoprecipitation of CD0386 from *C. difficile* 630

When anti-CD0386 IgG coupled resin was incubated overnight at 4°C with recombinant CD0386, the protein was successfully recovered, indicating that the coupling reaction was successful and that coupled anti-CD0386 antibodies were presented in a functional manner. However, when the resin was incubated with *C. difficile* 630 wall fraction, no protein was recovered (Figure 5.3.11)

5.3.8 Purification of *C. difficile* 630 Peptidoglycan

Despite three extended washes of insoluble cell debris with a solution of 4% SDS at 95ºC, SDS-PAGE revealed significant and persistent contamination with residual protein (Figure 5.3.12).

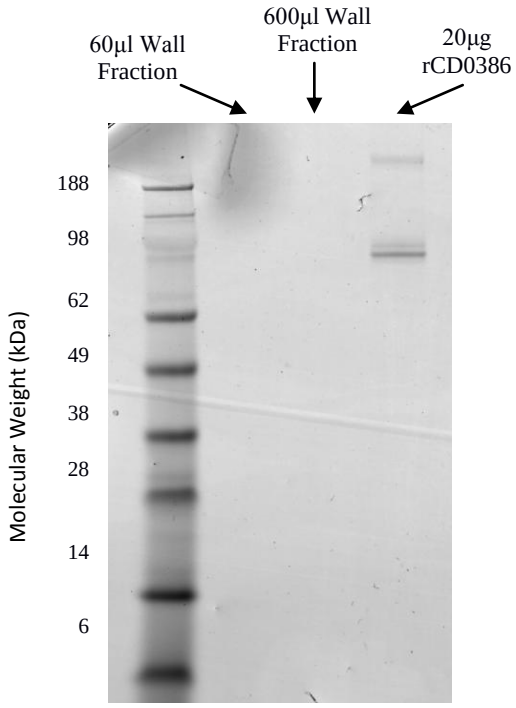


Figure 5.3.11 – Attempted Immunoprecipitation of CD0386 from *C. difficile* 630. Cell Wall fractions or recombinant CD0386 was incubated overnight at 4ºC with a resin conjugated to anti-CD0386 antibodies. Proteins were eluted by lowering pH and analysed by SDS-PAGE.

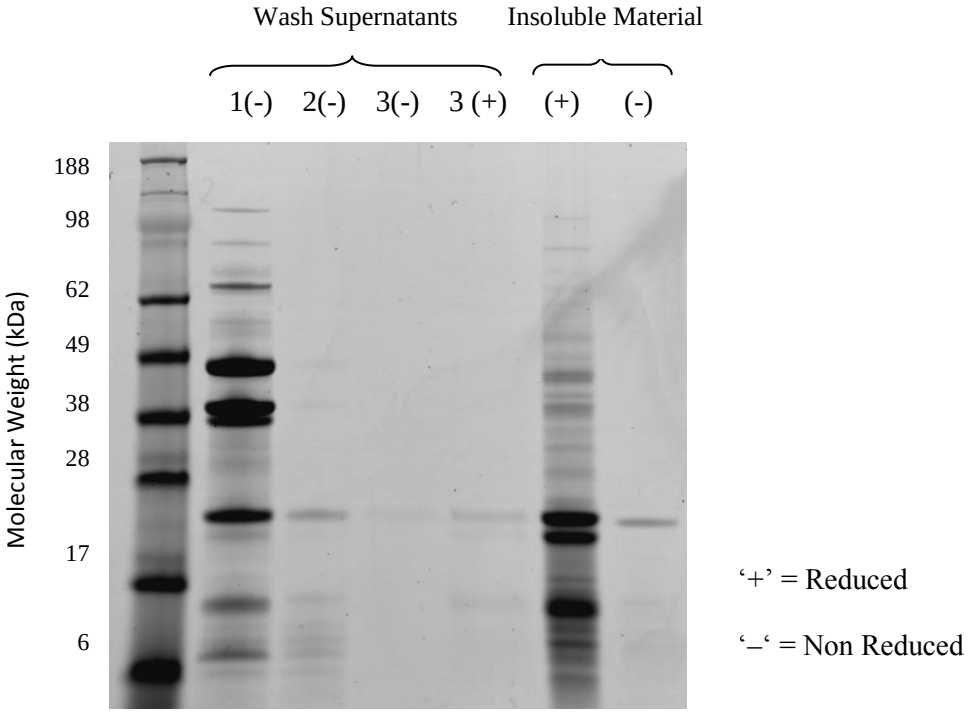


Figure 5.3.12 – Purification of *C. difficile* Peptidoglycan. SDS-PAGE analysis of supernatants following centrifugation of insoluble matter after washing with detergent. Insoluble material was resuspended in equal volume of loading buffer for analysis.

5.3.9 – Challenge of Hamsters with *C. difficile* 630 *srtB::erm*

Syrian hamsters were challenged with either *C. difficile* 630 or 630 *srtB::erm* as described in Section 5.2.10. Mortality data is presented in Figure 5.3.13

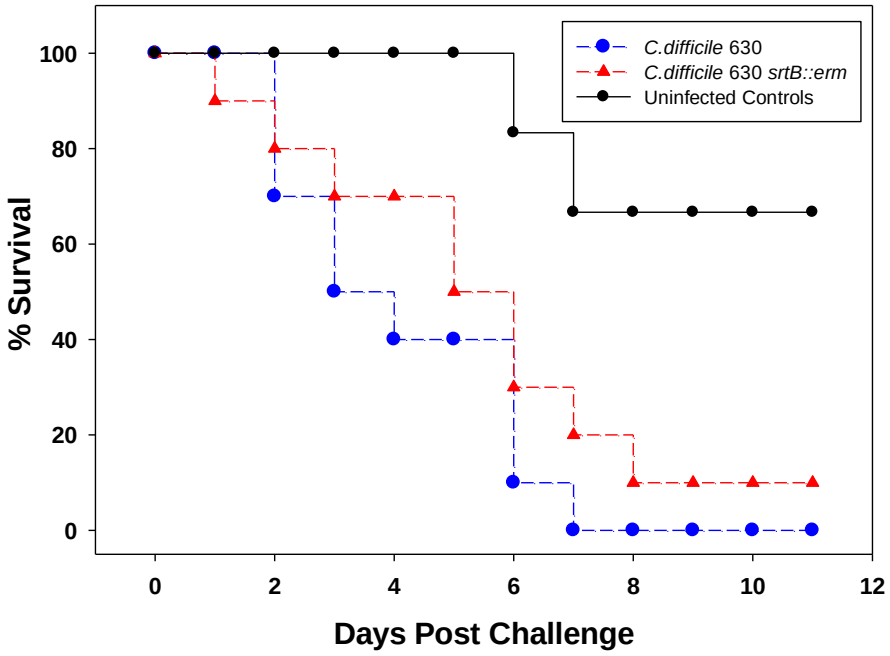


Figure 5.3.13 – Kaplan Meier Plot of Hamster Challenge with *C. difficile* 630 and *C. difficile* 630 *srtB::erm*. Plotted points indicate the day on which hamsters were euthanized due to level 3 symptomatic scoring.

Analysis by the Log-Rank test did not find any statistically significant difference between the mortality curves of hamsters infected with *C. difficile* 630 and *C. difficile 630 srtB::erm* ($p = 0.238$).

5.4 Discussion

5.4.1 - Introduction

Since sortase enzymes have no function other than to anchor other proteins to the cell surface, the phenotype of a sortase mutant is determined entirely by the functions of the cell wall anchored proteins.

The first aim of investigating a sortase mutant must, therefore, be to determine which surface proteins are anchored to the cell surface by the sortase and to confirm that their anchoring is absent in the sortase deficient strain. Following this, the wider phenotype of the sortase deficient organism can be investigated with the knowledge of the functions of its substrate proteins. In this chapter, the cellular distribution of one putative sortase substrate, CD0386 was examined in a sortase deficient mutant.

5.4.2 *In silico* Analysis of the *C. difficile* ORF CD2718 (Putative Sortase B)

Phylogenetic analysis of the open reading frame CD2718 (*srtB*) encoding the putative *C. difficile* sortase places it within the SrtB class. Of the small panel of species chosen for analysis, *C. perfringens* SrtB was the closest relative of CD2718.

Alignment of the translated *C. difficile* CD2718 with a selection of class B sortases reveals several conserved motifs. The most notable of these is the TLxTC motif near the C-terminus which is highly characteristic of sortase enzymes and contains the equivalent of the catalytic Cys184 in *S. aureus* SrtA. A conserved arginine is observed 8-10 amino acids towards the C-terminal of this motif, the equivalent of Arg233 in *S. aureus* SrtA. The intervening sequence forms the loop β7-β8 which is known to be variable in length in other species and indeed the alignment contains a gap of two amino acids in the *C. difficile* and *C. perfringens* sequences. Also of interest is the 'YGH' motif at position 114, in which the histidine residue is the conserved equivalent of His120 in the *S. aureus* SrtA. The functions of these conserved residues is described in detail in Section 2.3.

Considering the above observations, it appears that the *C. difficile* open reading frame CD2718 *(srtB)* encodes a complete sortase enzyme which most likely is of the class B.

5.4.3 The Putative Sortase Substrates and their Putative Functions

Based upon the possession of CWSS-like characteristics including an SPxTG or PPxTG sorting motif as proposed by Pallen et al (2001), seven putative *C. difficile* putative substrates were identified - 6 SPxTG proteins and one PPxTG protein. The set of proteins identified in this work concurs with the seven proteins identified by Comfort and Clubb (2004, supplementary data).

The proteins identified as putative sortase substrates consist of two putative cell wall hydrolases (CD0183 and CD2768), a putative 5' nucleotidase (CD2537), one protein of unknown function (CD3246) and three proteins (CD0386, CD2831, CD3382) with sequence homology with the *S. aureus* collagen binding protein, Cna and secondary structure homology with pilins from various species. In light of the classification of the *C. difficile* sortase as a class B enzyme, it is noteworthy that the proteins identified as its putative substrates are typical of the sortase repertoire observed for class A sortases rather than class B. Notably absent are typical iron-associated class B sortase associated proteins such as haem binding proteins reflecting the specialised role of SrtB in iron homeostasis. Based upon the above list of proteins, it would appear that although the *C. difficile* sortase is of the class B, it may fulfil a general 'housekeeping' role more akin to a class A sortase.

Proteins CD0183 and CD2768 are putatively identified as peptidoglycan hydrolases. Such enzymes are essential for processes such as cell growth, remodelling and turnover of peptidoglycan and more specialised roles such as creation of space for secretion systems (Vollmer *et al.*, 2008). Localisation of peptidoglycan hydrolases to the cell surface is a common adaptation, presumably in order to allow efficient access to substrates - for example, the peptidoglycan hydrolase InlB of *L. monocytogenes* contains glycine tryptophan rich repeat regions which mediate interaction with the lipoteichoic acid of the cell wall (Jonquières *et al.*, 1999). Sortase anchoring of peptidoglycan hydrolases would simply represent another method of localisation.

Protein CD2753 is putatively identified as a 5'-nucelotidase. Dephosphorylation of nucleotides by this protein may simply form part of a nucleotide scavenging pathway, as *C. difficile* contains several genes annotated as encoding members of the NCS1 nucleotide transporter family. However, there is evidence in other species that surface-localised nucleotidases may be virulence factors in their own right. For example, the AdsA protein of *S. aureus* is a putative 5' nucleotidase which produces adenosine by cleavage of the adenosine monophosphate found at elevated concentrations at the site of infection. AdsA is essential for virulence in a mouse sepsis model, a phenomenon thought to be due to inhibition of neutrophil activity by adenosine (Thammavongsa *et al.*, 2009). A similar virulence associated nucelotidase has been partially investigated in *Streptococcus sanguis* (Fan *et al.*, 2012). In both of the above examples, the nucleotidases in question are sortase substrates, containing LPxTG cell wall sorting motifs.

With respect to virulence, the proteins of most interest are those which display homology to adhesins of other species, namely CD0386, CD2831 and CD3392. There is evidence from an *in vivo* model that these proteins are virulence associated - in a porcine ligated loop model, transcription of CD0386 is upregulated 1.9- and 2.5-fold at 8hrs and 12hrs respectively and

CD3392 and CD2831 are upregulated 2.7-fold and 2.4-fold respectively, but only at 12hrs (Scaria *et al*. 2011, supplementary data).

CD2831 and CD3382 possess conserved domains identified as 'Cna B' domains. Cna is a two domain collagen binding protein from *S. aureus* - domain A possesses collagen binding activity while the B domain forms a stalk which projects the binding domain away from the cell surface. These proteins therefore only share homology with the structural domain of a collagen binding protein and their annotation as 'collagen binding protein' may therefore be premature.

Secondary structure prediction and homology analysis of CD0386, CD2831 and CD3382 suggests that all three possess homology with various bacterial pilins. They do not, however, possess the 'YPKN' motifs which are characteristic of sortase-associated pili and act as acceptors in the transpeptidation reaction. It is also notable that in western blots of cell extracts, CD0386 appears monomeric (see Fig. 5.3.9) with a single band rather than the polymeric, ladder-like pattern expected for a covalently polymerised pilin. CD0386 may, therefore, possess a pilin-like domain, but it is monomeric *in vivo.*

Perhaps the most intriguing of all the predicted sortase substrates is CD3246. Sequence analysis by Phyre2 predicted a very low level of secondary structure, resulting in a failure to produce any reliable secondary structure homology results. Similarly, a BLAST search results in no proteins with significant homology outside of low complexity regions at the C and N termini. In spite of the lack of homologous proteins, some clues to the function of CD3246 have been uncovered by workers investigating riboswitches. Lee and co-workers (2010) discovered a self splicing ribozyme upstream from the start codon of the CD3246 open reading frame. By monitoring cleavage of radiolabelled transcript, they discovered that the riboswitch is regulated by the presence of cyclic-di-guanidine monophosphate (c-di-GMP). The native transcript contains a stem loop which is likely to sterically hinder ribosome access to the start codon and a ribosome binding site which is distant from the start codon - splicing of the transcript in the presence of c-di-GMP results in a much more favourable transcript by removing the stem loop and creating a ribosome binding site which is an optimum distance from the start codon (Lee *et al.*, 2010). The importance of c-di-GMP in regulation of bacterial virulence has only recently been recognised, and encompasses twitching motility in *Pseudomonas aeruginosa*, flagellar motility in *Salmonella* Typhimurium and *Vibrio cholera* and biofilm formation in all of the aforementioned species as well as *Yersina pestis* (Tamayo *et al.*, 2007). In these examples, c-di-GMP modulates virulence via direct effects on effector proteins rather than by the translational regulation described for CD3246.

5.4.4 Generation of *C. difficile* 630 *srtB*::*erm*

A total of five sites provided by the Sigma Targetron tool were chosen and retargeted introns synthesised (see Table 5.2.1) of which the successful knockout at site 213 was the last to be attempted. At all other sites, failure occurred at the integration stage of the protocol - after streaking of transconjugants on agar supplemented with erythromycin, no colonies were obtained. Genetic confirmation of gene interruption was obtained by sequencing of the intron/exon junctions.

Although examination of transcripts from the interrupted gene was not performed in this work, results previously obtained suggest that insertion of the intron abolishes all transcription of the target gene, as evidenced by a lack of mRNA for the targeted gene (Kirby, 2011). The lack of reactivity when fractions of *C. difficile* 630 *srtB:erm* are probed with anti-SrtB rabbit serum (Figure 5.3.4B) provides good evidence that all expression of SrtB has been entirely abolished. The successful generation of a *srtB* mutant knockout clearly contradicts previous reports of that sortase is essential, reports which were based solely on the failure of the authors to isolate mutants of the *srtB* gene (Wren, 2011).

The major disadvantage of intron mutagenesis for gene inactivation, aside from its low rate of success, is that it is an insertional rather than a deletional method, resulting in a large genomic insertion into the gene of interest and the persistent introduction of an antibiotic resistance gene. More satisfactory methods such as double crossover allele exchange techniques are commonly used in other species but were described in *C. difficile* only recently (Cartman *et al.*, 2012). Using this technique, whole portions of the chromosome can be replaced by homologous recombination leaving no trace of manipulation aside from a ‘watermark’ element. The region of homology is placed on a replication defective 'suicide' vector flanked by two selection markers, one positive and one negative. Two crossover events are required – the first crossover results in integration of the entire vector into the chromosome, at which point integrants can be isolated using the positive selection system. The second crossover liberates the vector backbone from the chromosome leaving just the DNA between the regions of homology – mutants which have undergone double crossover can be isolated by the application of the counterselection agent, which will kill remaining single crossover mutants still harbouring the negative selection gene.

Another method which avoids large insertions while providing the additional benefit of allowing control of the mutant phenotype is inducible antisense RNA, an example of which in *C. difficile* has been described (Fagan and Fairweather, 2011). In this system, a sequence encoding a RNA complementary to the mRNA of the gene of interest was placed on a plasmid under the control of

tetracycline promoters. In strains transformed with these plasmids, effective silencing of the relevant genes could be achieved by addition of a non-toxic tetracycline analog to the growth medium.

5.4.5 Growth and Toxin Production

C. difficile 630 *srtB::erm* does not exhibit a growth deficit compared to the wild type. It grows at a similar rate and to a similar optical density during a logarithmic growth and maintains a similar optical density during stationary phase. This suggests that the *C. difficile* sortase does not have a role in cell nutrition - where sortases are involved in fulfilling nutritional requirements such as in the case of the iron-responsive SrtB enzymes, sortase knockout can lead to attenuated growth (Maresso *et al.*, 2006).

C. difficile 630 *srtB::erm* does, however, exhibit an aberrant toxin production profile. Intracellular TcdA concentrations were higher and significantly different ($p < 0.05$) at 7h, 24h and 50h. Mean extracellular TcdA concentrations were also higher, although this was only statistically significant at 24hrs. Altered toxin secretion is not necessarily an expected characteristic of a *C. difficile* sortase mutant, as none of the predicted sortase substrates have predicted functions related to toxin secretion, either directly or indirectly.

However, previous surface protein knockout strains have exhibited similar characteristics. In previous investigation of surface proteins utilising the ClosTron system (Kirby, 2011), increased toxin production was noted in 6 of 8 surface protein mutants. As increased toxin production appears to be a common artefact in surface protein mutants and therefore likely to be due to non-specific factors, it was not investigated further in this instance. A speculative explanation of the phenomenon relates to the toxin secretion apparatus - the classical *C. difficile* toxin secretion system is poorly characterised and may involve pores formed by the holin-like protein TcdE, (Govind and Dupuy, 2012; Olling *et al.*, 2012). Even subtle changes at the cell surface may disrupt this system leading to aberrant toxin expression or more generally affect the permeability of the cell envelope.

5.4.6 Localisation of Putative Sortase Substrate CD0386 in *C. difficile* 630 *srtB::erm*

To obtain cell fractions representing the cell wall, cell membrane and cytoplasm, a fractionation method was employed based upon digestion of the peptidoglycan by muramidases. Equivalent fractions collected from the two strains did not exhibit gross differences in protein profile, indicating that the fractionation method was similarly effective in both strains. As expected, most of the visible proteins in the wall fraction are SLPs (eg. the HMW and LMW products of SlpA cleavage), released as the peptidoglycan to which they are bound is digested by lysozyme and mutanolysin. As expected, cytoplasm and membrane fractions contained much greater numbers of proteins.

Wall, membrane and cytoplasm fractions of *C. difficile* 630 and *C. difficile* 630 *srtB::erm* were separated by SDS-PAGE, blotted onto membranes and probed with polyclonal rabbit serum raised against a recombinant version of the putative adhesin and putative sortase substrate CD0386. Probing with anti-CD0386 rabbit serum revealed that in wild type cells, CD0386 is present in the wall, membrane and cytoplasmic fractions in roughly equal amounts. In *C. difficile srtB::erm*, however, there is a dramatic loss of CD0386 from the cell wall fraction. This would be the expected effect of sortase inactivation if CD0386 is a sortase substrate. It is unlikely that this observation is due to changes in CD0386 expression since the relative proportion of CD0386 in the cytoplasmic fractions of the two strains is unchanged, and is unlikely to be due to differences in loading masses because protein concentrations were controlled by BCA protein assay to ensure that equivalent masses of total protein were loaded.

A faint immunoreactive band at a molecular weight slightly higher than that of the major reactive species is present in both the wildtype and sortase knockout strains. Given that cleavage of the cell wall sorting signal results in a loss in mass of 3.7kDa corresponding to the hydrophobic region downstream of the scissile bond, it is tempting to speculate whether this higher band corresponds to unprocessed CD0386. Unfortunately, as detailed in Section 5.4.7, preparation of native CD0386 for mass spectroscopy was not successful, so this hypothesis could not be tested. Fluorescence imaging of cells stained with anti-CD0386 rabbit serum followed by a fluorescent anti-rabbit conjugate indicates that CD0386 is expressed by only 1-3% of wild type *C. difficile* 630 cells at any one time. Counting the proportion of cells which were fluorescent did not find a statistically significant difference between the number of fluorescent cells in *C. difficile* 630 and *C. difficile* 630 *srtB::erm*. Further investigation into the phenomenon of the variable expression of CD0386 would be worthwhile to determine whether its expression is an example of true phase variation as is seen in the *C. difficile* surface protein CwpV (Reynolds *et al.*, 2011), or whether its expression is associated with a particular phase of growth.

In summary, the results presented above indicate that CD0386 is associated with the peptidoglycan of the cell wall in *C. difficile*, and that this association is lost in a sortase deficient mutant. It does, however, remain surface localised and can be visualised at the surface by immunofluorescence. There are several literature reports of sortase wall-anchored proteins which remain surface localised even in the absence of sortase activity. For example, Aucher et al (2011) investigated the sorting signals of several *B. anthracis* wall anchored proteins by fusing their C termini with GamR, a receptor for the *B. anthracis* gamma phage. Cells lacking SrtA were still susceptible to phage infection, indicating that GamR was surface associated even in the absence of sortase activity. In this case, it appears that the association, most likely via the hydrophobic region, was sufficient to render GamR functional and facilitate phage infection.

Evidence in other species suggests that even when surface localised sortase substrates are present at the cell surface, they may not be functional – deletion of *srtA* in *S. gordonii* results in a 97% reduction in binding to salivary agglutinin despite robust levels of non-anchored LPxTG-containing adhesins SspA and SspB detectable at the surface (Nobbs *et al.*, 2007). In this case, the surface presentation of the sortase substrate occurred but was clearly inappropriate for its function.

5.4.7 Immunoprecipitation

Attempts were made to immunoprecipitate CD0386 from cell extracts using anti-CD0386 antibodies with the intention to perform mass spectroscopy and determine whether a mass change was observed consistent with cleavage of the putative CWSS and loss of the hydrophobic region downstream of it – unfortunately sufficient yields were not obtained even from large volumes of culture, presumably due the low levels of expression of CD0386.

5.4.8 Preparation of Purified *C. difficile* Peptidoglycan

C. difficile peptidoglycan was prepared with the intention of digesting it with muramidases and hence producing a stringent fraction of proteins covalently attached to the peptidoglycan. This subset of proteins could then be analysed by proteomic methods and comparisons made between the wild type and the sortase mutant. The purification method employed is based upon the principle that proteins, DNA and lipid membranes (ie. the majority of cell components) are all soluble in a strong detergent solution, while SDS-insoluble by so can be recovered from SDS-solubilised lysate by centrifugation. Despite extensive optimisation of conditions and washing cycles, significant protein contamination of the peptidoglycan was observed. The SDS-PAGE band pattern of the contaminants is similar to that of fractions collected early during the procedure, suggesting that proteins are trapped within the peptidoglycan matrix and released upon preparation for SDS-PAGE. As such, purity of the peptidoglycan was judged as unsatisfactory for the intended purpose.

5.4.9 – Challenge of Hamsters with *C. difficile* 630 *srtB::erm*

The Syrian hamster is the longest established and most widely used animal model for *C. difficile* infection (Best *et al.*, 2012). Hamsters are exquisitely sensitive to *C. difficile* and typically die several days after infection. However, while diseased hamsters exhibit many of the histopathological features observed in human infection (Goulding *et al.*, 2009), they do not exhibit profuse watery diarrhoea that occurs in humans, and the rapid progression to fulminant disease is a poor representation of the usual progression of disease in humans. The extreme sensitivity of the hamster model may paradoxically render it insensitive to more subtle differences in colonisation and pathogenicity factors which may have real effects on the disease in humans.

Recent advances have seen the development of mouse models of CDI. Several groups have described a mouse model which models human disease more faithfully than the hamster model, with animals devloping diarrhoea and fulminant disease only occurring after challenge with high doses (Chen *et al.*, 2008). Relapse models have also been developed, closely mimicking the relapsing phoenomenon in humans (Sun *et al.*, 2011). It is also possible to induce a 'carrier' state in mice, whereby animals are persistently but asymptomatically infected, which again is a close mimic of human infection and has the potential to offer insight into how the disease spreads, how infection is established, and how outbreaks can be prevented (Lawley *et al.*, 2009). While it is likely that hamster models will remain the gold standard animal model for the foreseeable future while mouse models are refined and validated, it is likely that mouse models will offer a more accurate model of human disease in the long term.

In *Bacillus anthracis*, the sortase enzymes SrtA and SrtB have been found to be required for efficient survival in macrophages *in vitro* (alveolar macrophages being the primary site of infection in pulmonary anthrax), but in mouse model of cutaneous anthrax, *srtA* deletion mutants are fully virulent (Gaspar *et al.*, 2005; Zink and Burns, 2005). This may reflect different contributions of the sortases to disease in the two modes of infection, and given that pulmonary anthrax is the most serious form of the disease in humans, it is unfortunate that the sortase mutant was not tested in a pulmonary model. Similarly, in *L. monocytogenes*, there is disparity between *in vivo* and *in vitro* evidence – deletion of the *srtA* gene results in a loss of anchoring of proteins known to be essential for virulence, but in a mouse model, virulence is only moderately attenuated (Garandeau *et al.*, 2002). In. *S. aureus*, srtA is unequivocally essential for full virulence in animal models (Mazmanian *et al.*, 2000).

When hamsters are challenged with *C. difficile* 630 *srtB::erm*, mortality is indistinguishable from hamsters challenged with the wild type *C. difficile* 630. Two of the six uninfected control animals became symptomatic after several days - this does not indicate that the hamsters were

endogenously infected, since this would have resulted in control animals becoming ill ~2 days before challenged animals. More likely it reflects accidental cross-contamination, especially as the two were housed in the same box. Importantly, partial failure of the control group does not affect interpretation of the test groups since it occurred late in the experiment, long after the test groups became symptomatic.

While taking into account the limitations of the model, the results of hamster challenge indicate that SrtB is not required for pathogenesis of *C. difficile* in hamsters.

5.4.10 Further Work

The lack of a genetically complemented *srtB* mutant is a significant weakness of the work presented in this chapter. This omission was initially due to difficulties obtaining a suitable vector for the purpose. A vector with suitable replicons, derived from pMTL9301 (Purdy *et al.*, 2002) was eventually sourced and a plan was developed to modify it for complementation. Approval for this procedure was granted from the local genetic modification safety committee and the necessary notification was made, but time constraints ultimately prevented the work from being carried out.

In order to definitively confirm CD0386 as a *C. difficile* sortase substrate, it is essential to provide evidence of a mass change as would be associated with the transpeptidation reaction, ie. a loss of the cell wall sorting signal and replacement with a peptidoglycan fragment. Furthermore, tandem mass spectroscopy of recovered CD0386 could allow elucidation of the exact structure of the peptidoglycan anchor, as has been performed in *B. anthracis* (Budzik *et al.*, 2008b). Recovery of sufficient CD0386 is likely to be major obstacle to this experiment, and although expression of CD0386 from a plasmid under the control of a strong promoter may provide sufficient yield, *C. difficile* 630 is a non-attenuated clinical strain and CD0386 is a putative virulence associated protein, so it unlikely that such a plan could be executed. A more satisfactory approach would be the overexpression of a recombinant marker protein as a surrogate for CD0386, by fusing the marker protein to the C-terminus of CD0386 and therefore determining if the C-terminal CWSS is sufficient for surface anchoring. Such an approach was taken in *S. aureus* by fusing the C-terminus of SpA to the Staphylococcal Enterotoxin B protein (Mazmanian *et al.*, 1999) and in *B. subtilis* by fusion of the CWSS of the sortase substrate YchS to either a β-lactamase reporter protein or green fluorescent protein (Liew *et al.*, 2011).

Regarding investigation of the remaining six sortase substrates, piecemeal analysis by systematic raising of antisera as described in this work is feasible, but time-consuming and expensive. A more serious disadvantage of this approach is the severe bias encountered, because only proteins

suggested by bioinformatic analysis are investigated, leading to the risk that important unidentified sortase substrates could be overlooked. A more rigourous, non-assumptive approach would be to compare the peptidoglycan-linked proteomes of the wild type and sortase knockout. Methods such as surface shaving or biotinylation which have been very successful in general cell surface proteomics (Hempel *et al.*, 2011) are not suited to analysis of sortase mutants because, as has been discussed, sortase substrates often remain surface localised (albeit often non-functional) even in the absence of anchoring to the peptidoglycan. The focus of efforts should, therefore, be on developing methods based upon the digestion of peptidoglycan to isolate a stringent 'peptidoglycan-anchored' proteome.

5.5 Conclusions

Following *in silico* analysis of the putative *C. difficile* sortase and its putative substrates, a sortase deficient mutant of *C. difficile* 630 was generated by insertion of a bacterial intron into the *srtB* gene. Inactivation of the gene was confirmed by DNA sequencing and western blotting – the mutant exhibits normal growth and slightly aberrant toxin production. One of the seven predicted sortase substrates, the putative adhesin CD0386 exhibits a loss of anchoring to the cell wall in the sortase mutant. Efforts to recover CD0386 for mass spectroscopic confirmation of its processing by the *C. difficile* sortase were unsuccessful due to its low levels of expression under standard culture conditions. Plans for additional work to further develop upon these results have also been outlined.

Chapter 6 – Structural Characterisation of the *C. difficile* Sortase Enzyme

6.1. Introduction

6.1.1 Methods for Determination of Macromolecular Structures

One of the major driving forces behind the rapidly expanding knowledge of biology is the ability to visualise biological molecules, allowing researchers to directly examine the molecular basis of biological processes. As of 2013, the international Protein Data Bank (PDB) contains over 93,000 biological structures of proteins and nucleic acids, with its rate of growth increasing year on year. The most commonly used techniques for determination of macromolecular structures are Nuclear Magnetic Resonance (NMR) spectroscopy, X-ray crystallography and Electron Microscopy (EM). Other techniques such as neutron diffraction and Small Angle X-Ray Scattering (SAXS) are more specialist in nature and account for very few published structures. Historically, NMR spectroscopy has been more suited to small molecules and smaller macromolecules, but its use for protein structure is increasing due to the development of improved techniques for larger proteins. Conversely, EM is more suited to very large macromolecules because its resolution is ultimately limited by the wavelength of electrons. Although X-ray crystallography remains the most popular methodology employed by structural biologists (see Figure 6.1.1), diversity of techniques has somewhat increased in recent decades and many proteins have now been characterised using multiple techniques, providing valuable depth to understanding of their structure and function.

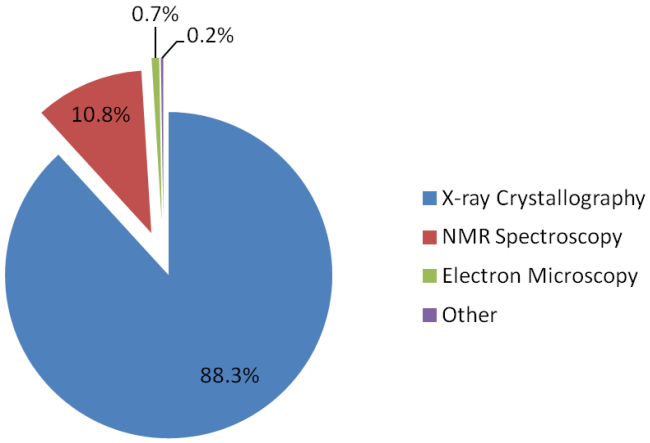


Figure 6.1.1 – Protein Data Bank Entries by Experimental Technique. Figures are shown as a percentage of PDB entries as of September 2013 The 'Other' classification includes Neutron diffraction, Powder Diffraction, Electron Crystallography and SAXS (www.rcsb.org/pdb/statistics/).

X-ray crystallography is suitable for characterisation of molecules over a wide range of molecular mass, from small molecules to complex proteinaceous assemblies such as viral capsids. It is also by far the most popular method of structure determination, with over 88% of the structures currently in

the PDB having been determined by this method. The major disadvantage of the technique is that it requires that the sample be crystallised, and because little is known about the mechanisms of crystallisation, conditions favourable for crystallisation must be determined empirically for each molecule, representing a major bottleneck in X-ray structure determination.

6.1.2 Crystallisation of Macromolecules

The first step in crystallisation is a nucleation event whereby two molecules in solution form a nucleus of crystallisation. Dependent upon the stability of these nuclei, growth of a crystal can then follow by ordered packing of subsequent molecules onto the nuclei. To encourage these processes, it is helpful to progressively increase the concentration of the molecule to be crystallised close its solubility limits to achieve a supersaturated state (see Figure 6.1.2). In protein crystallography, the most frequently utilised method to achieve supersaturation is vapour diffusion, whereby a drop is sealed within a chamber containing a reservoir of the crystallisation solution. Diffusion of water between the drop and the reservoir increases precipitant concentration in the drop, hence increasing the effective protein concentration. The drop may be placed on a pedestal within the sealed chamber ('sitting drop') or suspended over the reservoir by surface tension ('hanging drop').

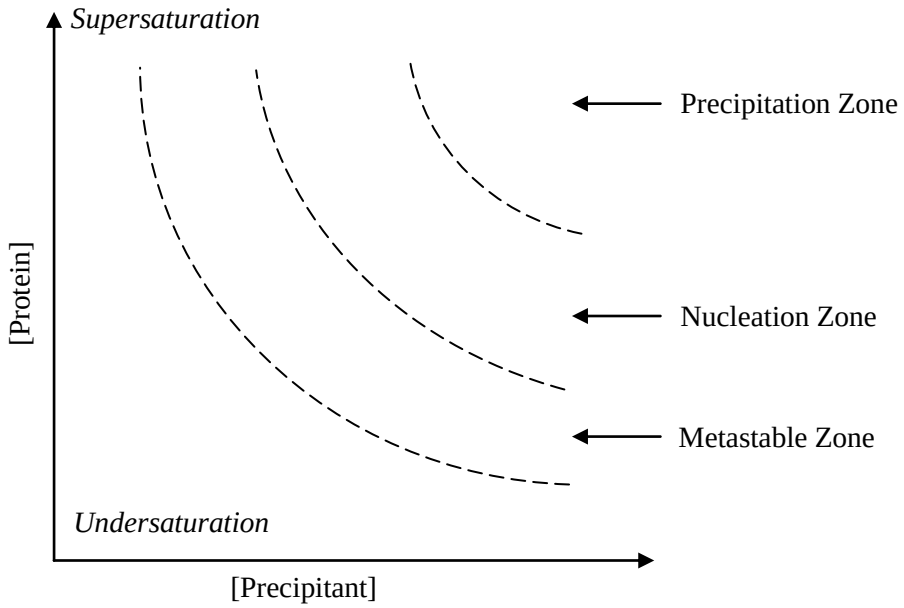


Figure 6.1.2 – A Hypothetical Crystallisation Phase Diagram. As the concentration of precipitant and/or protein concentration is increased, the state of the protein can pass through various several zones. Crystallisation experiments aim to place proteins in conditions beyond the metastable zone but without entering the precipitation zone.

To determine favourable conditions for crystallisation of a particular molecule, a large number of screening experiments must be performed, typically facilitated by high-throughput robotic systems which allow thousands of conditions to be tested rapidly and using a minimum of sample volume. Such screens frequently produce crystals of suitable size and quality for X-ray diffraction, but manual optimisation of conditions is generally required to optimal produce crystals.

Whilst nucleation is a prerequisite for crystal growth, the processes of nucleation and growth can be uncoupled experimentally so that crystals can be grown in conditions under which they do not readily nucleate (Chayen, 2005). The most common method to achieve this is known as 'seeding', whereby whole crystals or crystal fragments are transplanted into conditions more favourable for growth than for nucleation, often within the metastable zone. This can result in crystals of an entirely different morphology via epitaxial growth. Alternatively, seeding into conditions with a precipitant concentration lower than that required for spontaneous nucleation may yield larger crystals than would be obtained in the original condition.

6.1.3 The Properties of Crystals

Many crystalline substances display unusual physical properties, typically due to the arrangement of their constituent molecules in a highly ordered and regular manner. This regularity naturally gives rise to symmetry, and due to the number of symmetry operations possible in three dimensional space, there are a huge number of potential arrangements of molecules in a crystal – these arrangements are described in the 230 space groups listed in the Tables for Crystallography published by the International Union of Crystallography. With respect to protein crystals, the chirality inherent in proteins precludes mirror and inversion symmetry operations, reducing the number of potential space groups to 65. The majority of protein crystals fall into an even smaller set of the most common space groups.

6.1.4 X-Rays and Diffraction by Crystals

The unique usefulness of X-rays for molecular structure determination lies with their short wavelength which, at the lower range of 0.5 Å to 1.5 Å, overlaps the typical lengths of intra-atomic bonds (eg. 1.54Å for a C-C bond and 0.96Å for a O-H bond (Haynes and Lide, 2010)). Equally importantly, the interaction of X-rays with matter is sufficient to result in significant scattering of energy, while still readily passing through air.

The ordered crystal can be described in terms of a series of lattices. When a wave is diffracted by planes of atoms as exists in a crystal lattice, there are two potential outcomes. If the refracted waves are in phase with those from an adjacent lattice, the waves are said to interfere coherently. Bragg's law (see Figure 6.1.3) states that coherent diffraction occurs when the planes of diffracting

molecules are separated by a distance equal to an integer multiple of the wavelength. It is these peaks of coherent diffraction which are observed when a crystal is subjected to X-ray diffraction.

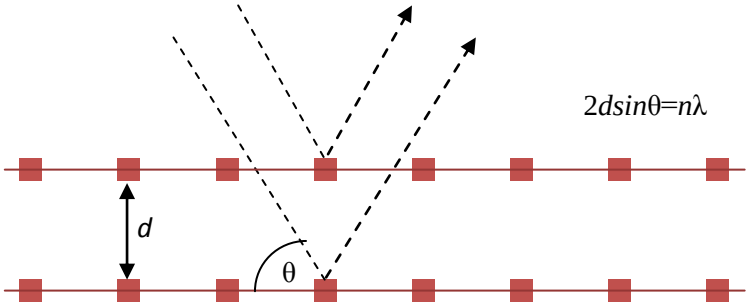


Figure 6.1.3. Illustration of Bragg's Law. X-rays diffract coherently where the distance between the two diffracting planes (*d*) is an integer multiple of the wavelength (λ). The additional path length of the lower wave is 2d*sin*θ.

The mathematical description of waves, especially complex waves such as those produced by X-ray diffraction, is often achieved by use of Fourier transforms. There are two forms of Fourier transform of interest to the crystallogapher. The forward Fourier transform is effectively a mathematical representation of the diffraction experiment, generating structure factors (amplitude and phase) from the electron density in the crystal. The reverse Fourier transform reconstructs the electron density in the crystal based upon the phase and amplitude of the waves. It is this operation which allows crystal structures to be determined from diffraction data. However, phases must be known - this is the so-called 'phase problem' which poses a major obstacle to structure determination as will be described in the next section.

6.1.5 The Phase Problem

The intensities of diffracted X-rays are easily recorded by use of photographic film or a charge coupled detector (CCD), but it is not possible to record the phase of the waves. This is a major obstacle to structure solution, since extracting information encoded in waves by Fourier analysis requires their phase to be known.

For many years, the predominant methodology for determining phases was Multiple Isomorphous Replacement (MIR), developed by Max Perutz during the 1950s (Perutz *et al.*, 1960). Compounds containing heavy metal atoms such as mercury or uranium are soaked into crystals where they bind at well-defined sites within the protein, such as thiol side chains in the case of mercury or carboxylic side chains in the case of uranium. Phase information is calculated by comparison of Patterson difference maps of the native and isomorphous heavy-atom soaked derivitaves.

The method of Multiple-wavelength Anomalous Dispersion (MAD), developed by Wayne Hendrickson and Jerome Karle, has become increasingly popular due partly to the widespread availability of tuneable wavelength beamlines at synchrotron facilities (Hendrickson *et al.*, 1988). This method, along with its predecessor Single-wavelength Anomalous Dispersion (SAD), takes advantage of the phenomenon of anomalous scattering at wavelengths close to the absorbance edges of atoms. Conveniently, the absorbance edge wavelengths of several heavy metals are close to the typical wavelengths used for protein crystallography. MAD does not require comparison of diffraction from *apo* and heavy metal derivative crystals – diffraction is instead recorded from the same crystal at different wavelengths. Selenium is a popular heavy atom for MAD because it can be consistently and accurately incorporated into proteins by recombinant expression in a methionine auxotroph organism grown in a selenomethionine substituted medium (Hendrickson *et al.*, 1990).

In addition to the experimental phasing techniques described above, several wholly computational methods are available. Increasing computing power has enabled the use of direct methods for structure solution using ‘Shake and Bake’ algorithms (Weeks *et al.*, 1999), but use of such techniques is still limited to data of atomic resolution. The most widely used purely computational phasing method for data of moderate resolution is ‘molecular replacement’ (Rossman and Blow, 1962), a technique which infers phase information from a previously solved homologous structure. The search model is placed within the unit cell and subjected to three-dimensional rotation and translation functions. A Patterson map is calculated from the test solution after each manipulation and compared to the experimental map. Convergence between the two maps will indicate a potentially correct solution from which phases can be calculated.

6.1.6 Refinement of Protein Structures

The aim of structure refinement is to improve the molecular model initially obtained by structure solution, doing so in such a way that the refined model better fits the experimental data. The fit between experimental structure factor amplitudes ($\boldsymbol{F}_{obs}$) and those predicted by the model ($\boldsymbol{F}_{calc}$) is assessed during refinement using statistical measures such as residual maximum likelihood or least square residual.

The simplest form of model refinement is rigid body refinement, whereby a whole model is moved rigidly – this is often performed at the very beginning of refinement to correct gross inaccuracies in the initial solution, following which more sophisticated techniques can be applied, whereby individual atoms or groups of atoms are moved in three dimensions. In most cases, refinement is restrained using known physical parameters such as bond lengths and angles. Not only does this ensure that the resultant structures are geometrically plausible, it increases the observation to parameter ratio and reduces computational demands. Unrestrained refinement is only appropriate when data is at or near atomic resolution, where experimental observations are of sufficiently high quality to produce a model of excellent geometry without restraint.

Residual based refinement has a disadvantage of risking the model becoming trapped in local minima – to overcome this, simulated annealing steps can be introduced, whereby the molecule is 'heated' to high temperature and slowly cooled. As the simulation is cooled, convergence between $\boldsymbol{F}_{obs}$ and $\boldsymbol{F}_{calc}$ is obtained with lower risk of the model being distorted by entrapment in local likelihood minima. Where non-crystallographic symmetry is present, restraints based upon NCS operators can be used, which again increases the observation to parameter ratio and can lead to significant improvements in the model.

6.1.7 Validation of Protein Structures

Several statistics are available to monitor the progress of refinement and to validate finished models. The most widely used is the crystallographic residual or, R_{cryst}, calculated by comparing an inverse reverse Fourier transform of the model to the experimental electron density map, thus providing a measure of how well the model explains the observed density.

However, reliance on R_{cryst} can lead to serious errors, because the atoms used for validation are also part of the refinement process, leading to a cyclical process whereby low R factors are easily achievable by over-fitting of the model to the data. This problem was addressed by the development of R_{Free}, a statistic which is independent of refinement, being calculated using a subset of reflections, generally 5-10%, set aside before refinement (Brünger, 1992). While some

difference between R_{cryst} and R_{free} is expected, especially at low resolution, inappropriate overfitting of a structure will produce an unusually large gap.

The geometric correctness of the model must also be assessed. A common method is to calculate root mean square deviation (RMSD) values for parameters such as bond lengths and angles. Local adherence to the ideal geometry of peptides can be assessed by examination of a Ramachandran plot, a graphical representation of the dihedral angles of carbon atoms in the peptide backbone (Ramachandran and Sasisekharan, 1968). Discrete areas within the plot represent the secondary structure of the peptide, encompassing regions of 'preferred' and 'allowable' combinations of phi and psi angle – amino acids which fall out of these regions are either incorrect or require justification by recourse to clear electron density.

Programs such as Molprobity (Davis *et al.*, 2007) perform more advanced geometric validation by assessing such parameters as dihedral angles, torsion angles and Cβ deviations alongside comprehensive analysis of van der Waal's contacts and hydrogen bonding. Ultimately, human interpretation of electron density is essential throughout model building and refinement to determine the biological plausibility of the model and its compatibility with the experimental observations.

6.2. Methods

6.2.1 Protein Purification

C. difficile SrtB and SrtB C226A were purified as described in Section 4.2.1.

6.2.2 High-Throughput Crystallisation Screening

Initial crystallisation conditions for *C. difficile* SrtB C226A were determined by high-throughput screening using a Phoenix robotic crystallisation system (Art Robbins Industries). Reservoir volumes of 50µl were dispensed into Intelliplate crystallisation plates (Art Robbins Industries) and sitting drops of 200nl assembled by dispensing protein and reservoir solution in ratios of 1:2, 1:1 and 2:1. Plates were sealed, incubated at 16°C and drops inspected microscopically weekly for one month, then monthly. Seven commercially available screens were tested - Structure Screen I and II, Clear Strategy Screen I, Clear Strategy Screen II, PACT Premier, Heavy and Light, JCSGplus and Morpheus (all obtained from Molecular Dimensions) .

6.2.3 Optimisation of Crystallisation Conditions

Manual optimisation of the condition 0.1M Phos/cit buffer pH4.2, 40% PEG300 was performed by the hanging drop method in a 16 well tissue culture plate with a reservoir volume of 1mL, sealed with greased 20mm siliconised glass cover slips. Variables optimised were pH of crystallisation solution (pH4-9), protein concentration (11.6-2.9mg/ml), PEG300 concentration (5-40%) drop volume (2-10µL) and ratio of protein to reservoir (1:3-3:1).

Microseeding was performed in parallel with optimisation of protein and precipitant concentration. A 1µL drop of small needles grown in the initial conditions was pipetted from its cover slip and resuspended in 50µl of 50% PEG 300, 100mM Phos-Cit pH7.2. This seed stock was then diluted in mother liqor in a range of 1×10^{-1} to 1×10^{-5} and 1µL added to fully assembled drops which had been allowed to equilibrate for 1hr.

Attempted crystallisation of SrtB under reducing conditions was performed as above but with the addition of *tris*(2-carboxyethyl)phosphine (TCEP) to protein stock and reservoir solutions at a concentration of 1mM. Attempted crystallisation of covalently inhibited SrtB was performed as above but with the addition of (2-(trimethylammonium)ethyl methanethiosulfonate (MTSET) to protein stock solutions at a concentration of 1mM.

6.2.4 Diffraction Data Collection and Processing

X-ray diffraction experiments with SrtB C226A crystals were performed at beamline I-04 of the Diamond Light Source, Didcot, Oxon. and diffraction recorded using a Pilatus 2M CCD detector.

Due to the 40% PEG300 concentration in the crystallisation conditions, no additional cryoprotectant was added prior to diffraction data collectin. Crystals were mounted within a cryoloop directly prior to data collection and frozen to 100K by placement within liquid nitrogen stream. Diffraction images were processed in space group $P2_12_12_1$ using the Xia2 pipeline (Winter *et al.*, 2013) at the Diamond Light Source and merged using Scala, a component of the CCP4 software suite (Winn *et al.*, 2011).

6.2.5 Structure Solution, Refinement and Validation

Initial phases were obtained by molecular replacement with Phaser-MR (McCoy *et al.*, 2007) using a homology model of *C. difficile* SrtB. The search model was produced by SWISS-Modeller (Arnold *et al.*, 2006) using the X-ray structure of *B. anthracis* sortase B (Zong *et al.*, 2004b) which exhibits 37% sequence identity and 63% sequence similarity to *C. difficile* SrtB. Refinement was performed using REFMAC (Murshudov *et al.*, 1997), each round of refinement consisting of three cycles of restrained refinement (XYZ, real space, B-factors, occupancies and torsion angle NCS) followed by manual refinement and rebuilding within Coot where required (Emsley *et al.*, 2010). The quality of the structure was evaluated throughout using Molprobity (Davis *et al.*, 2007).

6.2.6 Modelling of The Cysteine 226 Residue

In order to appropriately model the conformation of a modelled Cys226 residue, a superposition was assembled consisting of *C. difficile* SrtB C226A, *S. aureus* SrtB, and *B. anthracis* SrtB (Zhang *et al.*, 2004). The alanine 226 residue was mutated to cysteine and the chi (χ^1) angles of the mutated residue were adjusted to match the conformation of the template structures.

6.2.7 Structural Analysis

Superpositions were performed using the Secondary Structure Matching utility of Coot. Secondary structure assignments were performed using the *Define Secondary Structure of Proteins* (DSSP) program (Kabsch and Sander, 1983)

To identify structural homologues and analyse their relationship to the SrtB C226A structure, the PDB was queried using the PDBeFold program hosted by the European Bioinformatics Institute (EBI) (Krissinel and Henrick, 2004). Calculation of protein-protein interfaces and assembly surface area was performed using PDBePISA, also hosted by the EBI (Krissinel and Henrick, 2007).

All figures were prepared and rendered using Pymol (Schrödinger LLC).

6.3 Results

6.3.1 Crystallisation of SrtB C226A

Initial high-throughput screening experiments yielded clusters of needle crystals (Figure 6.3.1 A) in the condition C6 of the JCSGplus HT96 screen (0.1M Phosphate/Citrate pH4.2, 40% PEG300). Optimisation of protein concentration and precipitant concentration yielded thicker crystals (Figure 5.3.1 B), with the optimum parameters being a drop volume of 2µl, a protein stock to reservoir ratio of 1:1, 40-34% PEG300 and a protein concentration of 2.9mg/mL. Even larger crystals (Figure 6.3.1 C) were obtained by microseeding, the optimum parameters being 32% PEG300, a protein concentration of 2.5mg/mL and a seed dilution factor of 1 $x10^{-5}$.

Crystallisation of wild type SrtB was not successful, either by high throughput screening or optimisation of conditions under which SrtB C226A was crystallised, including the addition of TCEP as a reducing agent and the irreversible sortase inhibitor MTSET.

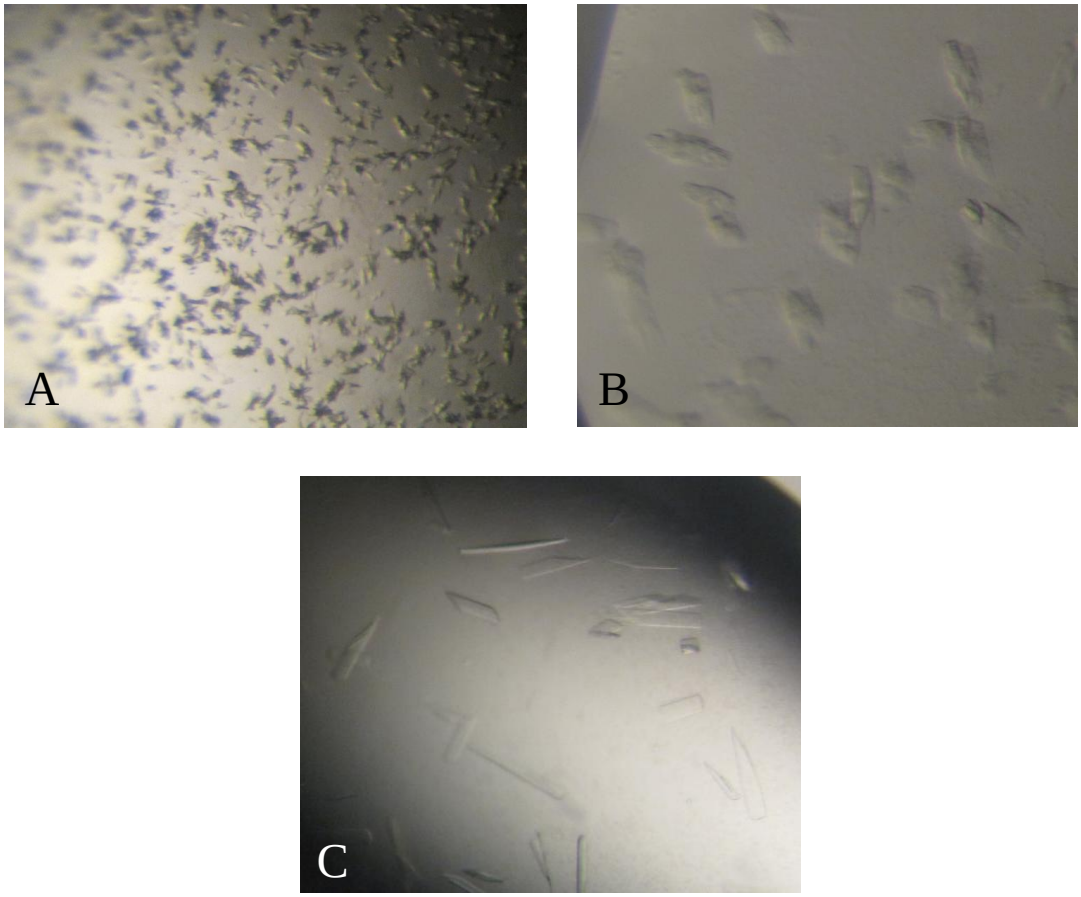


Figure 6.3.1 – Crystallisation of SrtB C226A. (A) Crystals grown in 0.1M Phos/Cit buffer pH4.2, 40% PEG300 with 10mg/ml SrtBC226A (B) Larger crystals grown after optimisation of protein concentration to 2.9mg/ml (C) Crystals grown by microseeding into pre-equilibrated drops containing 2.5mg/ml SrtB C226A, 0.1M phosphate/citrate pH4.2 and 32% PEG300.

6.3.2 X-Ray Diffraction with SrtB C226A Crystals

Diffraction of X-rays with crystals of *C. difficile* SrtB was performed on 3rd May 2012 at beamline I-04 of the Diamond Light Source, Harwell Innovation Campus, Oxfordshire, UK.

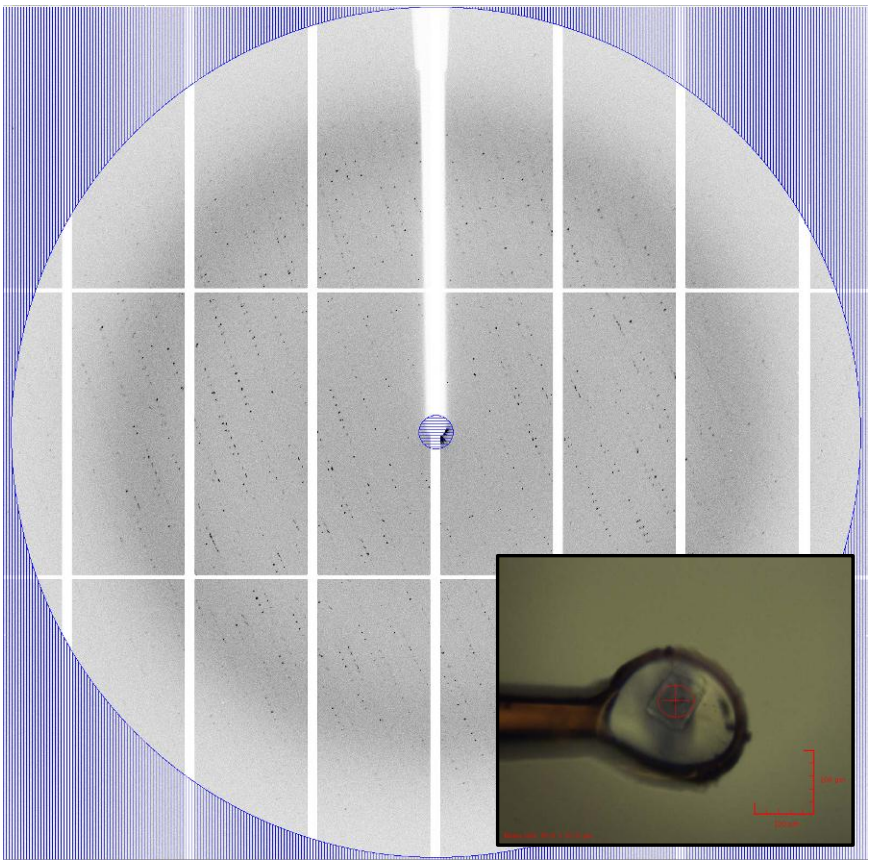


Figure 6.3.2 – X-Ray Diffraction by Crystals of *C. difficile* SrtB C226A. An example diffraction image is shown with a diffraction limit of 2.5Å, represented by the blue hatched area. Inset is an image of the mounted crystal at the beamline.

Three sets of diffraction images were collected, a summary of which is shown in Table 5.3.1. A summary of data collection statistics is displayed in Table 6.3.2.

Dataset Name	No. of Images	Phi range	Resolution limits (High, low)	Completeness (%)
Cd2718_1_M3S5_1	60	0-60	2.31, 29.41	66.0
Cd2718_1_M3S5_2	120	140-260	2.47, 29.45	90.2
Cd2718_1_M3S5_3	100	260-360	2.28, 29.13	84.6

Table 6.3.1 – Summary of images collected by diffraction of SrtB C226A

6.3.3 Diffraction Data Processing

Data reduction, indexing, integration and scaling in $P2_12_12_1$ was performed by the automated Xia2 pipeline (Winter *et al.*, 2013). Merging of the three datasets was later performed manually using Scala. Due to the low completeness of Set 1, only Sets 2 and 3 were merged. A resolution limit of 2.55Å was applied as the R_{merge} above this resolution was unacceptably high. Processing statistics for these datasets are shown in Table 6.3.2.

Data Collection and Processing Statistics	**SrtB C226A**
Space Group	$P2_12_12_1$
Number of protein molecules per asymmetric unit	2
Cell dimensions	a = 38.25Å, b = 90.27Å, c = 134.74Å $\alpha = \beta = \gamma = 90^o$
Resolution range (Å)	29.37 – 2.55
R_{symm} (outer shell)	0.151 (0.506)
I/σI (outer shell)	11.8 (4.2)
Completeness (outer shell) %	98.8 (100.0)
Total no. of reflections	122009
Unique no. of reflections	15902
Redundancy (outer shell)	7.7 (8.0)

Table 6.3.2 – Summary of Data Collection and Processing Statistics for SrtB C226A.

6.3.4 Solution of SrtB C226A Structure by Molecular Replacement

Determination of initial phases was performed by molecular replacement using Phaser-MR (McCoy *et al.*, 2007). The search model was a homology model of *C. difficile* SrtB based upon the *B. anthracis* SrtB, which exhibits a 37% sequence identity and 63% sequence similarity to *C. difficile* SrtB. Based upon the Matthews coefficient calculation, Phaser-MR predicted an asymmetric unit containing ~54kDa of protein, relative to SrtB theoretical mass of~28kDa. The program therefore searched for two molecules in the asymmetric unit. Molecular Replacement Statistics are shown in Table 6.3.3. The >4-fold increase in LLG upon the addition of the second molecule is strongly suggestive of a correct solution, as is the TFZ of >8.0.

	RFZ	TFZ	PAK	LLG
Molecule 1	6.8	7.8	0	75
Molecule 2	4.2	16.7	2	317

Table 6.3.3 – Molecular Replacement Statistics for SrtB C226A. Where RFZ = Rotation Function Z-Score, TFZ = Translation Function Z-Score, PAK = Packing Clashes, LLG = Log Likelihood Gain.

6.3.5 Refinement and Validation of SrtB C226A

Refinement was performed iteratively across the whole assembly using Refmac for automated refinement and Coot for real space refinement following inspection of electron density. The model exhibited severe discordance with the electron density in two regions – the helical regions from the N-terminus to β1 and the tight loop between β5 and β6. Here, the modelled amino acids were deleted and replaced with a polyalanine chain, allowing the main chain atoms to be fitted before addition and refinement of side chains. A summary of refinement and validation statistics is shown in Table 6.3.4.

Model Refinement Statistics	**SrtB C226A**
R_{cryst}/R_{free}	0.19/0.24
Average B-factor (Å^2)	
Overall	24.49
Water	25.57
Protein	24.47
Clashscore	3.95
RMSD	
Bond length (Å)	0.006
Bond angle (^{o})	0.998
Ramachandran statistics	
Favoured (%)	95.56%
Additionally allowed (%)	4.44%
($R_{cryst} = \Sigma_h \|F_o - F_c\|/\Sigma_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h, respectively. R_{free} is as for R_{cryst} for a randomly selected 5.0% subset of reflections not used in refinement)	

Table 6.3.4 – Summary of Refinement Statistics for SrtB C226A.

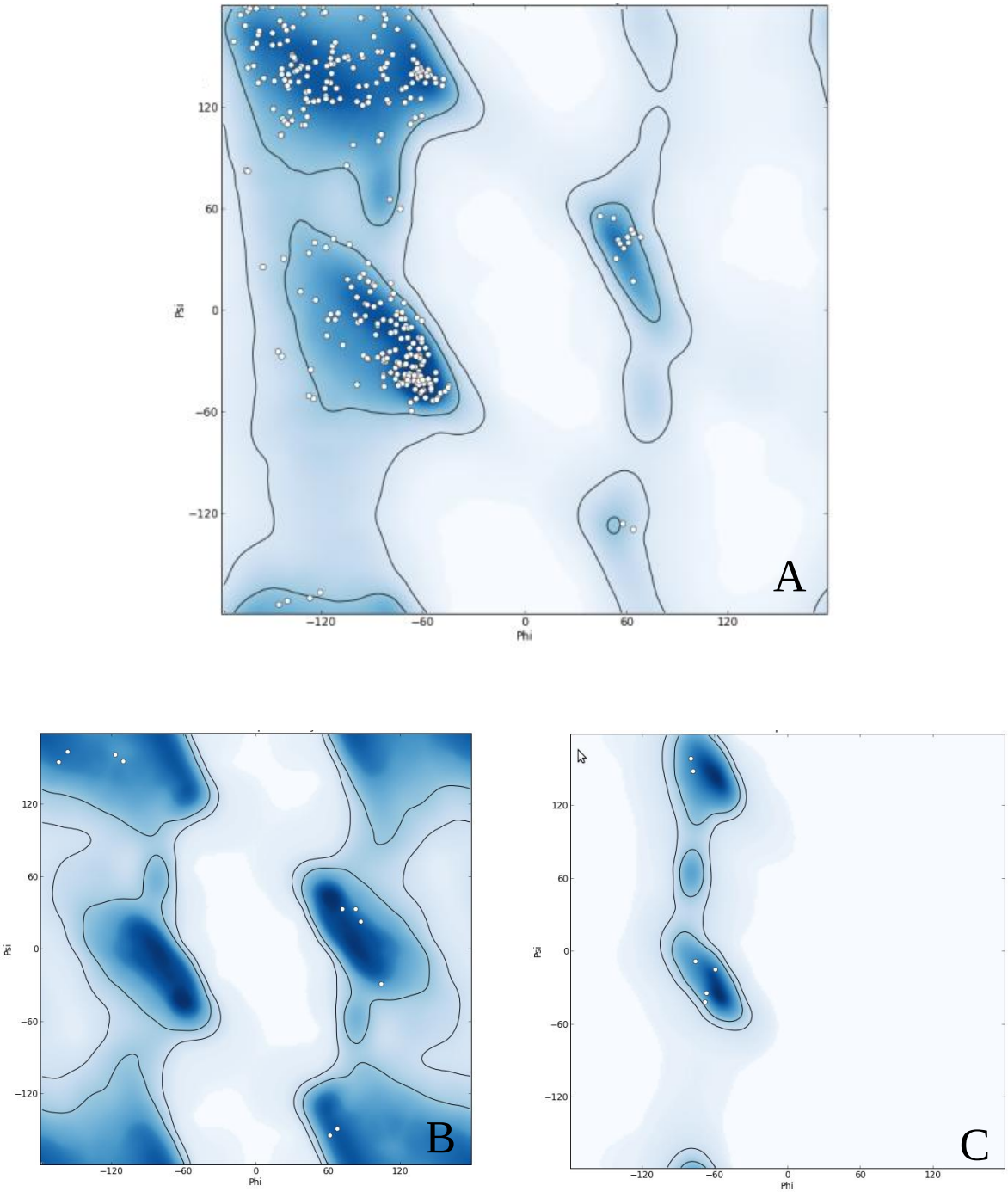


Figure 6.3.3 - Ramachandran Plots for SrtB C226A. Ramachandran plots of SrtB C226A generated for (A) All non-Proline/Glycine residues (B) Glycine residues and (C) trans-proline residues. Solid lines represent the limits of 'allowed' and 'preferred' angles.

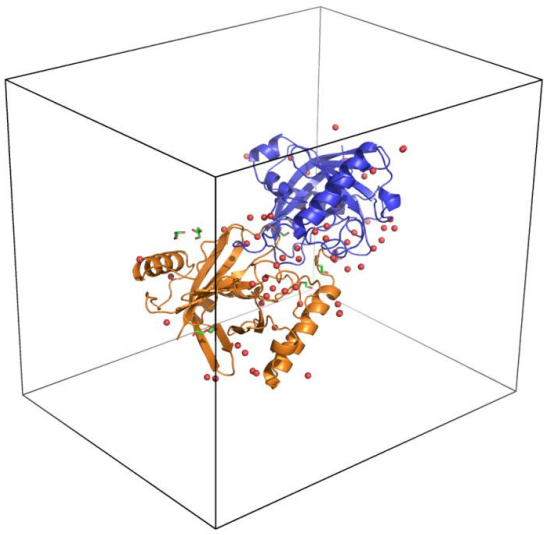
6.4. Analysis

6.4.1 Crystallisation of the *C. difficile* SrtB C226A and Solution of its Structure

The *C. difficile* sortase SrtB with a cysteine to alanine substitution at position 226 was crystallised in phosphate/citrate buffer at pH4.2 and with 40% PEG300, conditions determined by high throughput screening. Microseeding into the metastable zone (< 40% PEG300) was required to grow crystals of sufficient size for X-ray diffraction. Crystals diffracted X-rays to 2.28Å and data were processed in $P2_12_12_1$ with a 2.55Å resolution limit. The structure was solved by molecular replacement using a homology model based upon the *Bacillus anthracis* sortase B as a search model, the intention of homology modelling step being to use the most closely related extant structure while minimising model bias in the final structure.

The model of SrtB C226A contains two molecules of SrtB C226A, with the modelled dimer including 115 water molecules and 7 diethylene glycol molecules derived from the PEG300 present in the crystallisation solution. The two SrtB molecules exhibit a high level of similarity, and can be superposed with a Cα RMSD of 0.30Å.

Figure 6.4.1 - One SrtB C226A Crystallographic Dimer Shown Within the Unit Cell. The model is composed of two monomers of SrtB C226A (blue and orange) and includes 115 water molecules (red spheres) and 7 diethylene glycol molecules (green sticks).



SrtB C226A Chain A

SrtB C226A Chain B

The refinement statistics show in Table 6.3.4 are considered to be acceptable for a structure of this resolution, with an R_{cryst} of 0.19 and an R_{free} of 0.24. The $R_{cryst} - R_{free}$ interval of 0.05 is considered acceptable, although indicates a slight overfitting of the model. No Ramachandran outliers are present, and although 4.44% of residues are in non-favoured regions of the plot, this is for the entire dimer, so equates to only 2.22% per protein chain. Bond lengths and angles are also within the acceptable range for a structure of this resolution and analysis by Molprobity did not reveal any dihedral angle outliers or rotamer probability outliers.

6.4.2 The Structure of *C. difficile* SrtB

The arrangement of secondary structure elements and overall fold of *C. difficile* SrtB (Figures 6.4.2 and 6.4.3) is typical of a class B sortase. The N-terminus is characterised by a substantial alpha helix of 18aa followed by a short loop incorporating a three residue 3_{10} helix. Strands β1 and β2 are arranged in an antiparallel manner, separated by a turn motif. A long loop places β3 (which is the shortest of the beta sheet regions) parallel to β2, while β4 is antiparallel to β3.

Following β4, another 3_{10} helix of 7 residues is followed immediately by a shorter alpha helix of 4 residues. This breaks the continuity of the barrel by placing β5 adjacent to β1 rather than β4. The longest extended beta sheet region of 13aa, β6, lies antiparallel to β5 for around a third of its length and forms a striking curved backbone of the barrel. The long sequence separating β6 from β7 features an alpha helix of 12 residues in length. β7 runs parallel to β4, with β8 completing the barrel by running between and antiparallel to β6 and β7.

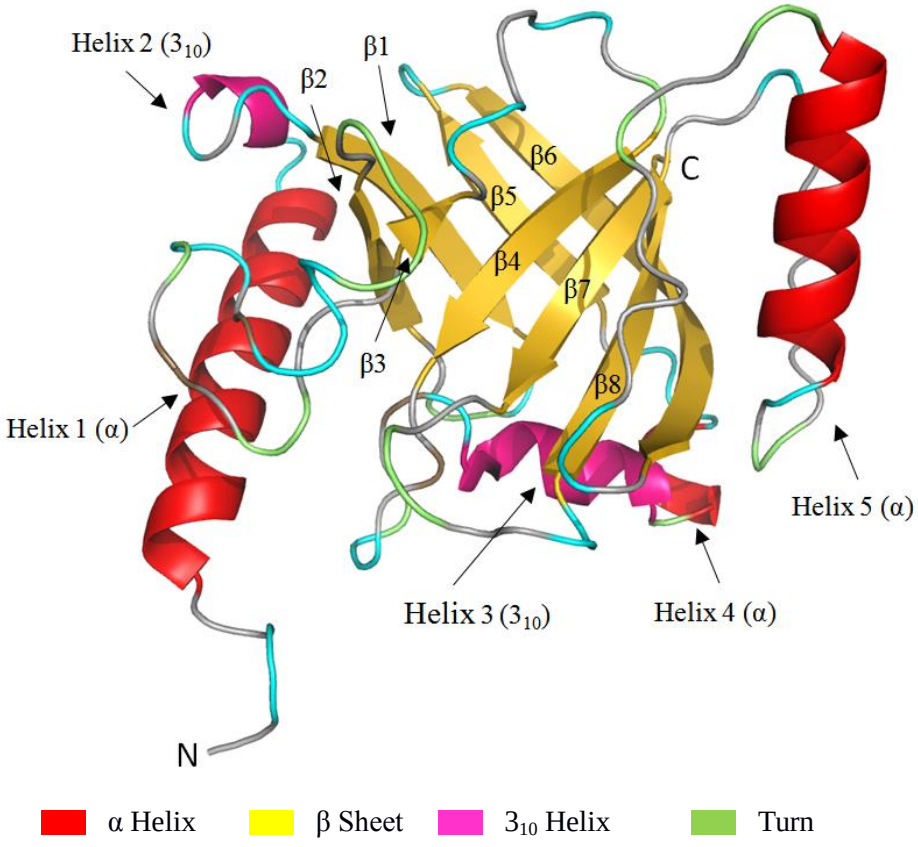


Figure 6.4.2 – *C. difficile* SrtB C226A Coloured by Secondary Structure. The structure of *C. difficile* SrtB incorporates three α helices, two 3_{10} helices and eight β strands. Secondary structure assignment was performed using DSSP.

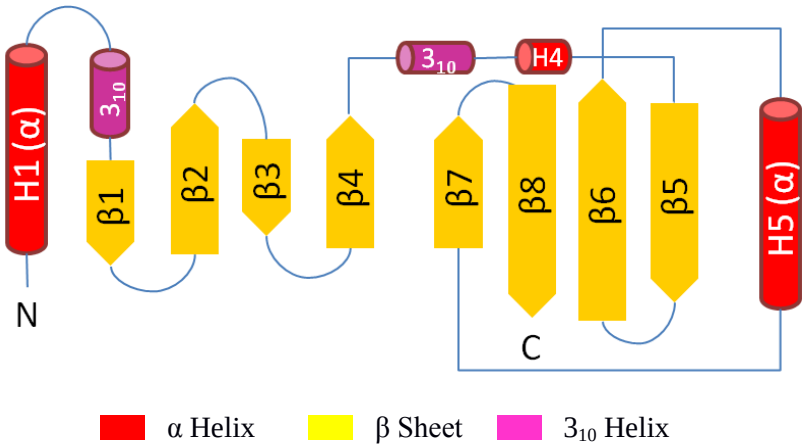


Figure 6.4.3 – Schematic of *C. difficile* SrtB C226A coloured by Secondary Structure. A schematic of secondary structure elements clearly shows the non-contiguous nature of the barrel from β4 onwards.

There are several regions of *C. difficile* SrtB C226A which exhibit B factors significantly higher than the structure average, indicating high mobility (Figure 6.4.4). These include the β3-β4 loop which has been implicated in metal binding in other species (see Section 6.4.5) and the β6-H5 loop which contains the putative substrate binding region of *C. difficile* SrtB (see Section 6.4.5)

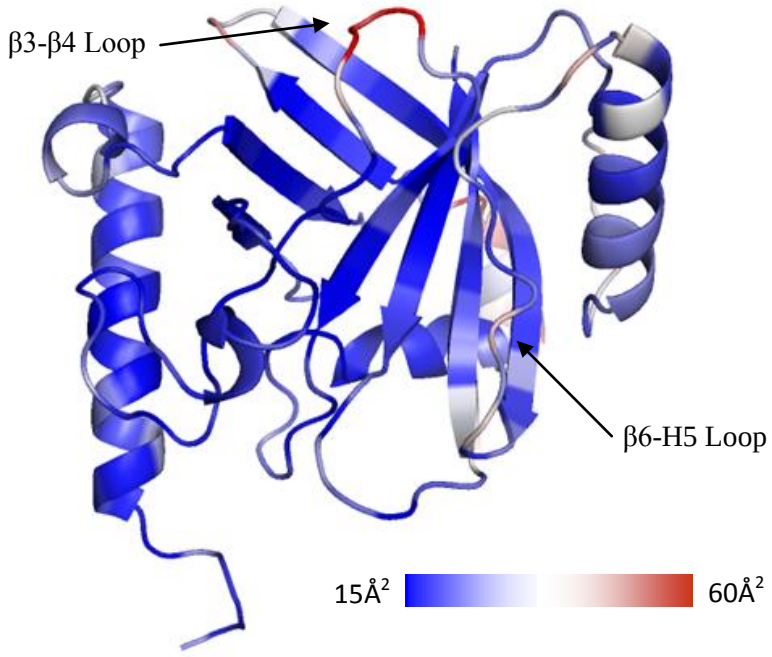


Figure 6.4.4 – Structure of *C. difficile* SrtB C226A coloured by B-Factor. B-factors higher than the structure average of 25Å^2 are observed in β3-β4, β5-β6 and β6-H5.

6.4.3 General Comparison of *C. difficile* SrtB with Other Sortase Enzyme Structures

To provide a comparison of *C. difficile* SrtB with structures of related proteins, secondary structure analysis and three-dimensional alignment was performed using the PDBeFold program, results of which are shown in Table 6.4.1. The structure of *B. anthracis* SrtB is the closest homologue, with the two other SrtB structures ranking second and third. Secondary structure matching scores were lower for the class A and C sortases although the RMSD of matched atoms was similar, presumably due to similarities in the barrel region which is conserved between the sortase classes.

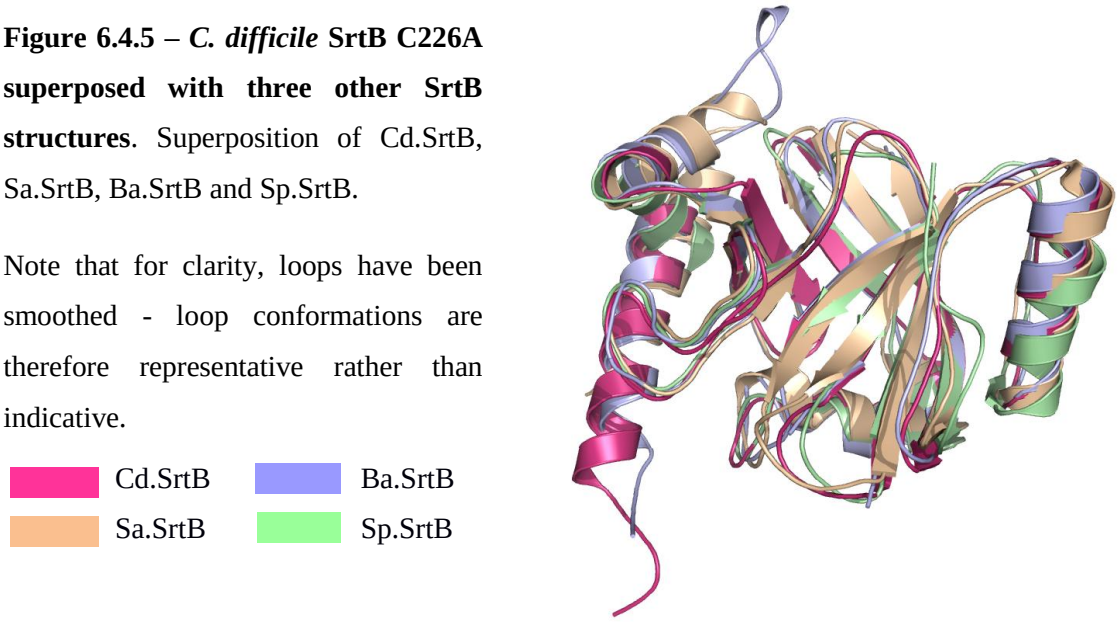
Class	Species	Cα RMSD (Å)	SSE* Match (%)	PDB Code	Reference
SrtB	*B. anthracis*	1.19	77	1RZ2	(Zhang *et al.*, 2004)
SrtB	*S. aureus*	1.41	85	1NG5	(Zhang *et al.*, 2004)
SrtB	*S. pyogenes*	1.62	77	3PSQ	N/A
SrtC	*A. oris*	1.88	62	2XWG	(Persson, 2011)
SrtA	*S. pyogenes*	1.97	54	3FN6	(Race *et al.*, 2009)
SrtA	*S. aureus*	1.99	54	1T2O	(Zong *et al.*, 2004a)
SrtC	*S. agalactiae*	2.00	62	4G1H	(Khare *et al.*, 2011)

Table 6.4.1 – PDBeFold Analysis of *C. difficile* SrtB C226A. Results from PDBeFold analysis, sorted by ascending RMSD (*SSE = 'Secondary Structure Element').

Figure 6.4.5 shows a superposition of *C. difficile* SrtB (Cd.SrtB) with those of *S. aureus*, *B. anthracis* and *Streptococcus pyogenes* (Sa.SrtB, Ba.SrtB and Sp.SrtB respectively). The barrel structure and topology is identical in all four SrtBs, as are the placements of helices H3-H5.

Figure 6.4.5 – *C. difficile* SrtB C226A superposed with three other SrtB structures. Superposition of Cd.SrtB, Sa.SrtB, Ba.SrtB and Sp.SrtB.

Note that for clarity, loops have been smoothed - loop conformations are therefore representative rather than indicative.



As shown in Figure 6.4.6, the most striking structural differences between the four molecules lie at their N-termini. In Sa.SrtB and Ba.SrtB, this region features two alpha helices (H1 and H2) arranged in an inverted 'V' and where H1 lies roughly parallel to the axis of the barrel. Ba.SrtB possesses 13aa loop between these helices which projects a three residue long 3_{10} helix away from the core structure of the enzyme, while in Sa.SrtB the helices are connected by a much shorter loop of 4 residues. The N-terminus of Sp.SrtB is different to all three, possessing a single helix at 90° to the axis of the barrel.

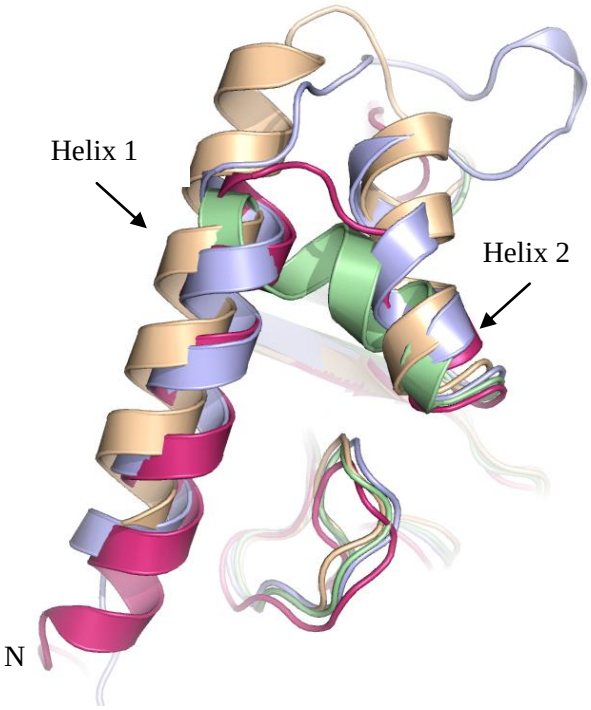


Figure 6.4.6 – N-terminus of *C. difficile* SrtB C226A superposed with three SrtB structures. The N-termini of *B. anthracis S. aureus, S. pyogenes* and *C. difficile* SrtB enzymes adopt three different conformations. Labelling is for *C. difficile* SrtB.

Cd.SrtB
Ba.SrtB
Sa.SrtB
Sp.SrtB

The structure of this region in Cd.SrtB is most similar to that of Sa.SrtB, but with a significantly shorter H2, resulting in H1 being placed lower relative to the side of the barrel. Reflecting their novel arrangement in Cd.SrtB, these two helices were placed incorrectly in the initial solution produced by molecular replacement and significant manual rebuilding of the entire N-terminus was required. Nothing is known of the functions of the N-terminal helices, which are a unique feature of the class B sortases, although it has been suggested that their purpose is to project the enzyme away from the cell surface (Zong *et al.*, 2004b). Differences in this region may reflect differences in bacterial physiology as well as function, which would explain the radically different conformation in the pilus-associated SrtB of *S. pyogenes*.

6.4.4 The *Clostridium difficile* SrtB Active Site

C. difficile SrtB features the characteristic conserved sortase triad of histidine, cysteine and arginine centred upon Cys226 at the C-terminal end of strand β7 (Figure 6.4.8). Cys226 and Arg224 reside in a deep depression on the surface of the enzyme, while the conserved His133 is partially shielded from the depression by the side chain of Tyr227 within the β7-β8 loop (Figure 6.4.7).

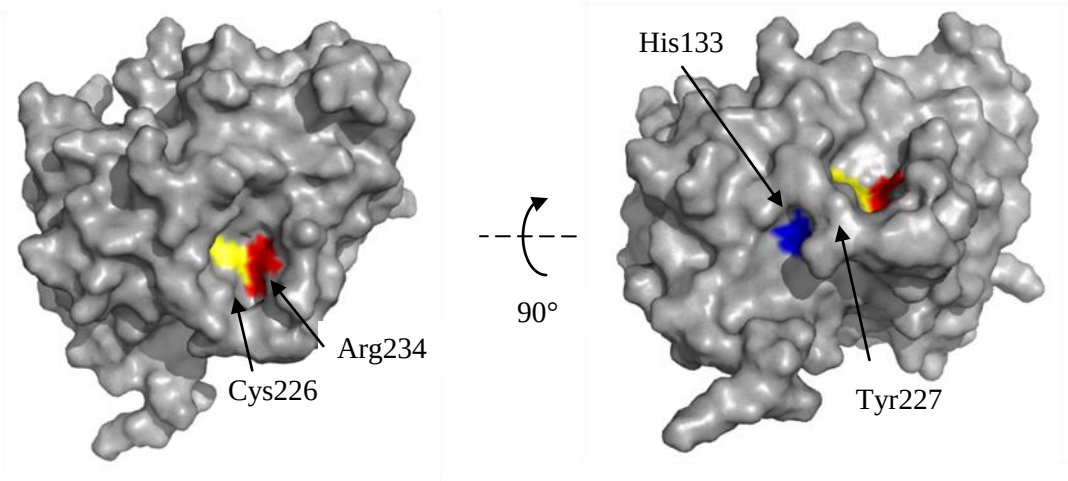


Figure 6.4.7 – Surface Representation of the *C. difficile* SrtB Active Site. The conserved cysteine and arginine residues form the floor of a deep depression at the C-terminus of β7. The adjacent conserved histidine at the base of β4 is also visible from behind loop β7-β8.

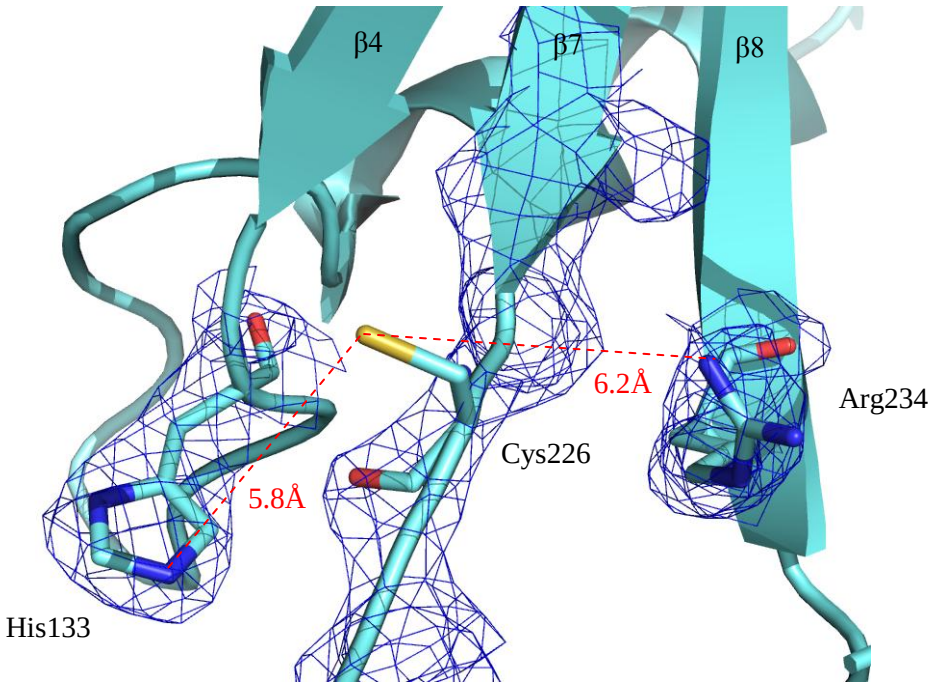


Figure 5.4.8 – The Modelled Active Site of *C. difficile* SrtB. Selected electron density (blue mesh) is shown contoured to 1.0 Sigma. Clear and continuous electron density is visible for His133 and Arg234. Electron density is clearly absent for the modelled Cys266.

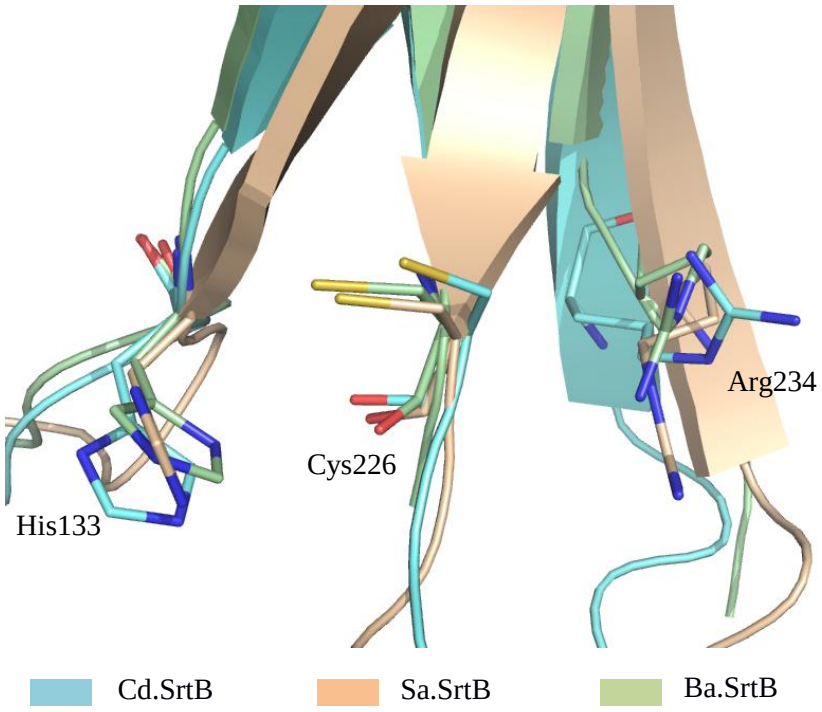


Figure 6.4.9 – Superposition of Three SrtB Active Sites. Superposition of the Cd.SrtB active site with Sa.SrtB and Ba.SrtB. Residue numbering is for Cd.SrtB.

The 5.8Å distance between Cys226 and His133 in *C. difficile* SrtB is greater than *B. anthracis* and *S. aureus* SrtB enzymes (4.2Å and 4.6Å respectively), but is similar to that observed in *S. aureus* SrtA (5.7Å). The increased distance relative to the other class B sortases is due partly to a slightly different histidine conformation but mostly due to a torsional displacement of the β7-β8 loop. It is not possible to exclude that distortion of the loop is an artefact of the Cys to Ala mutation.

Of the three SrtB active site residues, the conserved arginine appears the least strictly conserved with respect to its side chain conformation. In the *B. anthracis* structure it is clearly projected towards the conserved cysteine residue. In the *S. aureus* structure it points downwards, towards the β7-β8 loop. In the *C. difficile* SrtB it appears directed upwards, away from the loop. Nevertheless, the distance between Cys226 and Arg234 in Cd.SrtB is similar to that of Ba.SrtB due to aforementioned rotational displacement of the cysteine residue backbone.

6.4.5 The Putative CWSS Substrate Binding Site of *C. difficile* SrtB

As there are no published structures of a SrtB complexed with a Cell Wall Sorting Signal (CWSS) peptide, analysis of the putative substrate binding region of *C. difficile* SrtB must be based on comparison with the *S. aureus* SrtA-CWSS structure (Suree *et al.*, 2009). As described in Section 2.5.3 the β6-β7 loop of *S. aureus* SrtA undergoes structural rearrangements to produce a 3_{10} helix which contributes to an enveloping hydrophobic environment for the hydrophobic 'LP' portion of the LPxTG *S. aureus* SrtA CWSS. Such rearrangements are not thought to be universal to sortase enzymes, as the structures of several SrtA enzymes exhibit such a helix in *apo* form (See Figure 2.5.6). The presence of H5 in SrtB results in β6-H5 being analagous to the β6-β7 loop of Sa.SrtA.

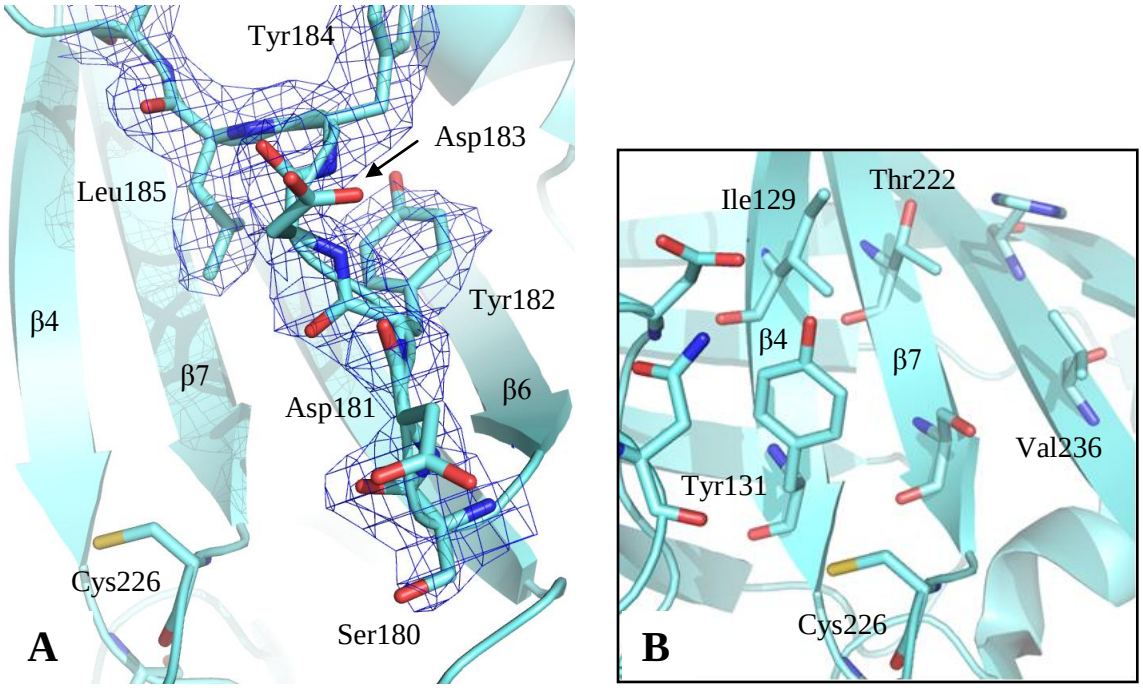


Figure 6.4.10 – The Putative CWSS Peptide Binding Loop of *C. difficile* SrtB. (A) The β6-H5 loop analogous to the peptide binding region of *S. aureus* SrtA contains polar residues with one conserved hydrophobic residue, Leu185. Selected electron density (blue mesh) is shown contoured to 1.0 Sigma. Examination of the floor of the binding cleft (B) reveals several residues (Ile129, Tyr131, Thr222 and Val226) which could provide hydrophobic contacts with the proline of the SPKTG motif.

In contrast to the exclusively hydrophobic interactions of Sa.SrtA with its LPxTG substrate, the Cd.SrtB, Sa.SrtB and Ba.SrtB sorting signals of SPKTG, NPQTN and NPKTG respectively require a more polar environment for the P1 residue - serine and asparagine are capable of acting as both donators *and* acceptors in hydrogen bonding interactions. The β6-H5 loop adjacent to the Cd.SrtB active site (Figure 6.4.10) is predominantly polar but also contains a semi-conserved hydrophobic

residue, Leu185. This residue and its Leu/Ile homologues in the other two SrtB enzymes may be analogous to Leu169 of Sa.SrtA which is located at the top of the 3_{10} helix and interacts hydrophobically with the P2 proline (see Fig 2.4.5 for comparison). Hydrophobic contacts with the invariant P2 proline are likely to be the similar in all sortase enzymes, and residues capable of such contacts are present in Cd.SrtB (Fig 6.4.10 B).

All three SrtB enzymes contain at least one tyrosine residue in the sequence preceding the conserved Ile/Leu (Figure 6.4.11 B), with Ba.SrtB and Cd.SrtB possessing two. Of particular interest is Cd.SrtB Tyr182 which is conserved in the primary sequence of Ba.SrtB, while in Sa.SrtB, the side chain of an adjacent tyrosine residue is placed only ~3Å away from that of Cd.SrtB (Figure 6.4.11 A) suggesting this may be a structurally conserved feature although further SrtB structures are required to confirm this. This residue is a good candidate to form a polar contact with the serine of the 'SPKTG' CWSS.

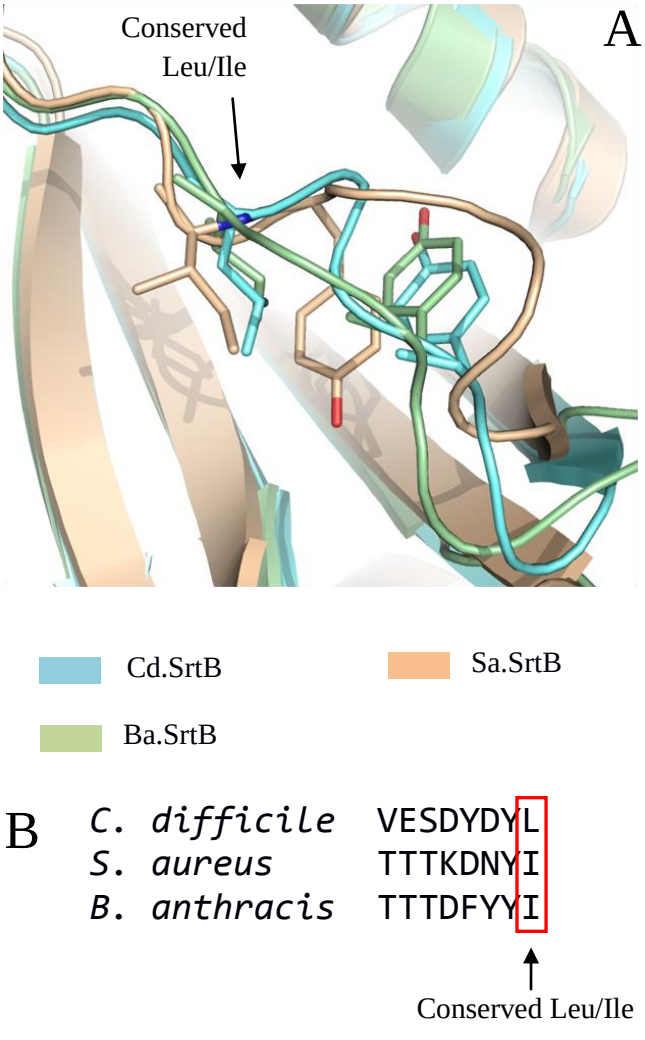


Figure 6.4.11 – The Putative CWSS Peptide Binding Loop of Three Superposed SrtB Enzymes. (A) The β6-H5 loop of *C. difficile* SrtB, *S. aureus* SrtB and *B. anthracis* SrtB all contain a conserved Leu/Ile residue and a structurally semi-conserved Tyrosine residue. (B) The conserved polar nature of the amino acids preceding the conserved Leu/Ile is clear in the primary sequence of the region.

6.4.6 Structure of the *C. difficile* SrtB Putative Transpeptidation Substrate Binding Site

The structure of *S. aureus* SrtB/GlyGlyGly revealed a binding site for the transpeptidation substrate within the β7-β8 loop (Zong *et al.*, 2004b), confirmed by NMR experiments in *S. aureus* SrtA (Suree *et al.*, 2009). The analagous region in *C. difficile* SrtB is shown in Figure 6.4.12.

The loop contains two prominent aromatic residues which project upwards towards the barrel and shield the sides of the area between the loop and the active site. Within this shielded area is the side chain of Arg233, which is thought to be responsible for activating the transpeptidation substrate for nucleophilic attack by abstraction of a proton from its N-terminus (See Section 2.5.4). Unlike in Sa.SrtB where the GlyGlyGly peptide hydrogen bonds with backbone as well as with the side chain of glutamic acid residue directly adjacent to the cysteine, no side Cd.SrtB chains project into the space between the loop and the active site. Any hydrogen bonding interactions between Cd.SrtB and its transpeptidation substrate must therefore involve backbone amide and carboxyl groups.

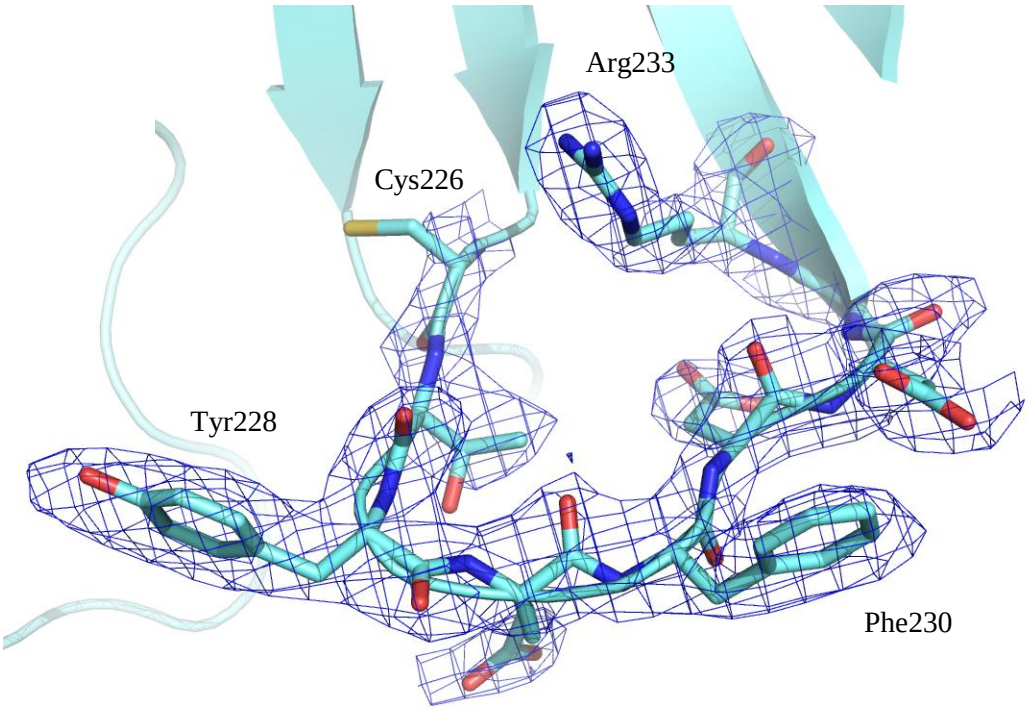


Figure 6.4.12 – The β7-β8 Loop of *C. difficile* SrtB C226. The loop contains two aromatic amino acids, the side chains of which shield the putative substrate binding region between the loop and the active site. The conserved arginine residue, responsible for activating the transpeptidation nucleophile, is also observed projecting into this space.

As measured from the conserved cysteine at the end of β7 to the conserved arginine within β8, Cd.SrtB features a loop of 7 amino acids, which is the same length as the Sp.SrtB but two amino acids shorter than those of Sa.SrtB and Ba.SrtB. The structural implications of this can be seen in Figure 6.4.13. The loop adopts a similar conformation in all three despite differences in sequence length, with an inward bend which is more subtle in Cd.SrtB.

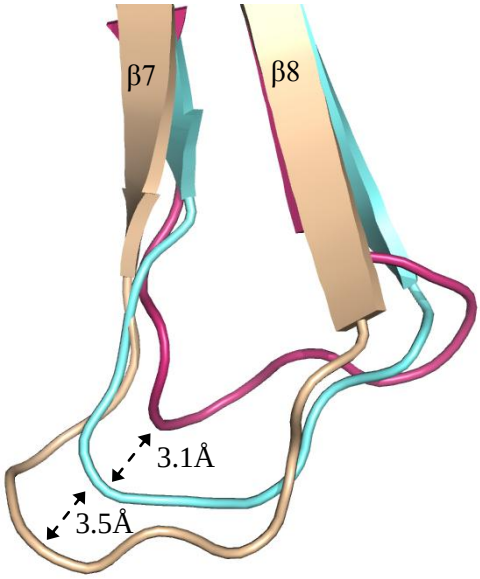


Figure 6.4.13 – β7-β8 Loop of *C. difficile* SrtB C226A Superposed with two SrtB Structures.

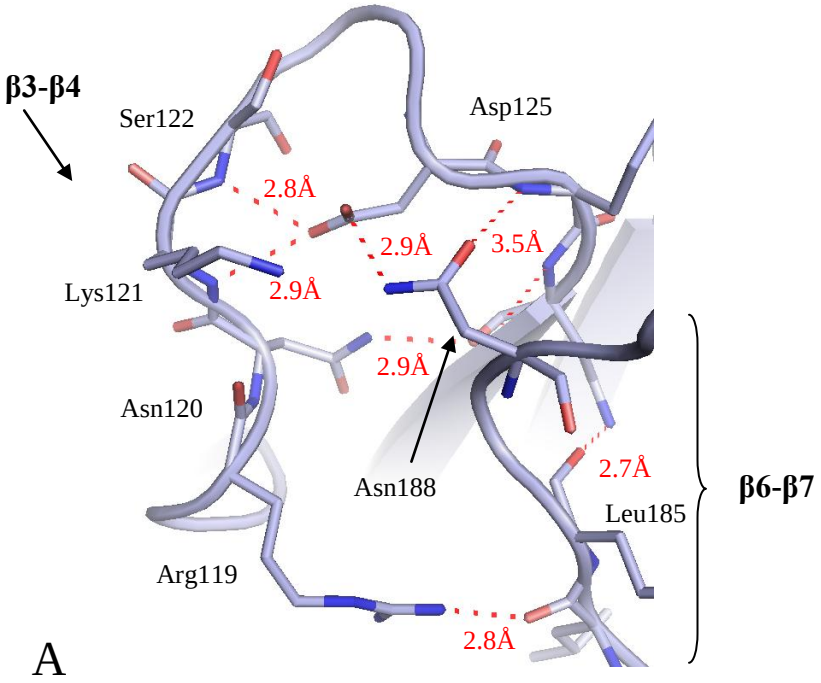
The β7-β8 loop of Cd.SrtB adopts a similar conformation to that of Sa.SrtB, but its shorter length results in a shorter extension, ~3.5Å closer to the barrel region relative to Sa.SrtB

Ba.SrtB β7-β8 was omitted from the superposition due to poor electron density.

Cd.SrtB
Sp.SrtB
Sa.SrtB

5.4.7 Structure of the *C. difficile* SrtB loop β3-β4

In *S. aureus* SrtA, which is dependent upon calcium for activity, a Ca^{2+} ion resides in a pocket formed by three glutamic acid residues in loop β3-β4 (Figure 6.4.14 B). The ion also engages a fourth glutamic acid within β6-β7, hence stabilising the substrate binding loop and thus exerting its stimulatory effect. In the *C.difficile* sortase, this region exhibits an extensive network of hydrogen bonds (Figure 6.4.11 A) which stabilise the loop without the requirement for metal ions.



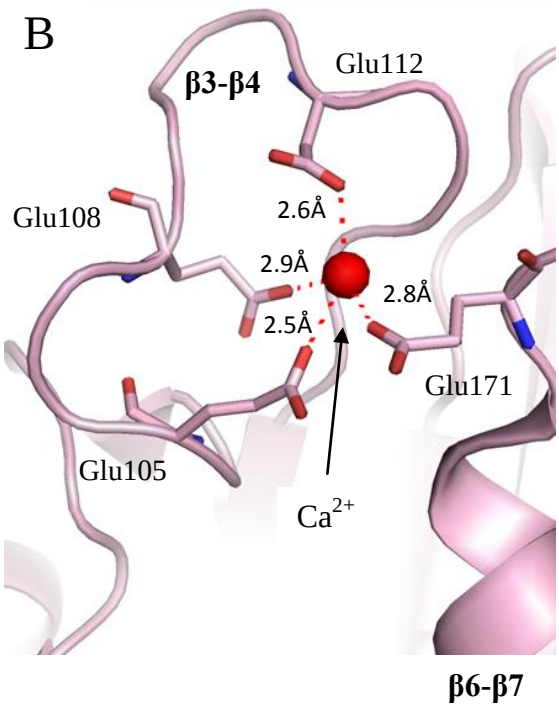


Figure 6.4.14 - Metal Independent Stabilisation of the β6-β7 Loop in *C. difficile* SrtB. (A) Distances consistent with hydrogen bonding are observed between Asn188 of loop β6-β7, and the backbone amide and side chain of Asp125 within loop β3-β4. The β3-β4 loop itself is stabilised by hydrogen bonding of Asp125 with backbone amides of Lys121 and Ser122. Two additional stabilising hydrogen bonds are observed between Leu185, Arg119 and Asn127. This is contrasted with the *S. aureus* SrtA (B) where stabilisation of β6-β7 is facilitated by Ca^{2+}.

6.4.8 Analysis of Non-Crystallographic Dimerisation of SrtB C226A

The non-crystallographic symmetry within SrtB C226A crystals results from dimerisation of two SrtB C226A molecules. As can be seen in Figure 6.4.15 A, these interactions are extensive in nature, with the N-terminus of each chain projecting into the active site depression of the other subunit (see Figure 6.4.2 B). These interactions are not thought to be biologically relevant, as size exclusion chromatography experiments indicate that SrtB is monomeric in solution. Crystal structure of sortase enzymes frequently exhibit crystallographic polymers, although no such structures exhibit interfaces similar to those observed in *C. difficile* SrtB.

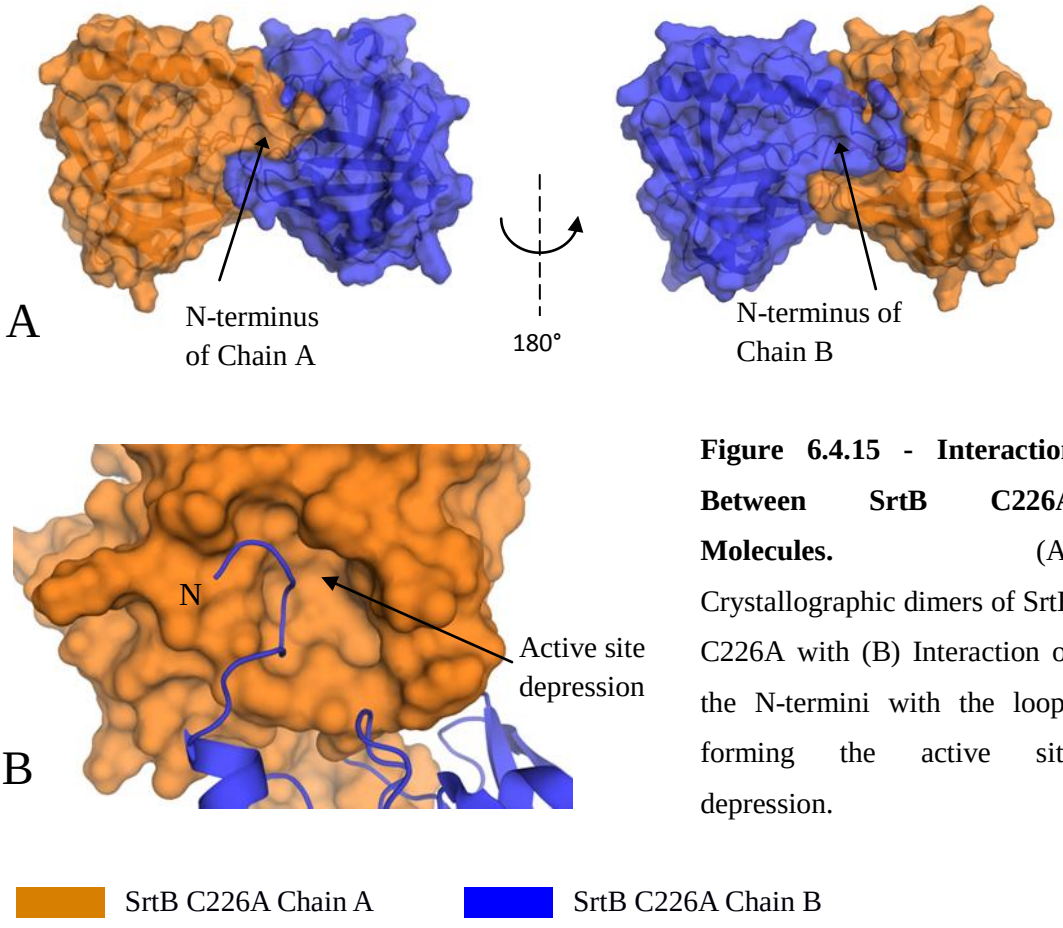


Figure 6.4.15 - Interaction Between SrtB C226A Molecules. (A) Crystallographic dimers of SrtB C226A with (B) Interaction of the N-termini with the loops forming the active site depression.

Analysis with PISA reveals 13 intermolecular hydrogen bonds. There are two interfaces: 1) Between the N-terminus and the loops surrounding the active site (β6-β7/β2-β3) and 2) Between loop β7-β8 and β6-H5. The assembly surface area is calculated to be 19580 $Å^2$, with a buried surface area of 4150 $Å^2$. The interface is not symmetrical – 5 of the 13 hydrogen bonds occur in only one chain. Tabulated results of the PISA analysis is shown in Table 6.4.2.

Bond	Distance (Å)	Chain A Residue [Atom]	Chain B Residue [Atom]
1	2.42	Asp 183[OD2]	Lys 40[HZ3]
2*	2.37	Asn 96[O]	Gly 42[H]
3	2.28	Ser 97[OG]	Gly 42[H]
4*	2.14	Lys 39[O]	Asn 96[HD22]
5*	1.88	Thr 227[O]	Lys 136[HZ2]
6*	2.04	Asn 134[O]	Phe 230[H]
7*	2.34	Gly 42[H]	Asn 96[O]
8	2.47	Ala 44[H]	Ser 97[OG]
9*	2.08	Asn 96[HD22]	Lys 39[O]
10	2.48	Lys 136[H]	Tyr 228[O]
11*	1.65	Lys 136[HZ3]	Thr 227[O]
12	3.37	Thr 227[OG1]	Glu 229[OE1]
13*	2.07	Phe 230[H]	Asn 134[O]

Table 6.4.2 – Hydrogen Bonds between SrtB C226A molecules. Entries with an asterisk indicate the presence an equivalent bond in the other chain.

6.5 - Discussion

6.5.1 General Discussion

The *C. difficile* sortase enzyme adopts a fold similar to other class B sortases, being composed of the 8-stranded non-contiguous β-barrel which forms the core of all sortase classes, decorated on its exterior surfaces with the multiple helices which are characteristic of the class B and C sortases. Compared to other class B enzymes, it features a more compact arrangement of helices at the N-terminus. Little is known about the function of these helices, but being located proximally to the membrane anchor region it is likely that they are involved in enzyme localisation and hence variation in this area is likely to reflect differences in either enzyme function or bacterial cell surface physiology.

By analogy with Sa.SrtA, the putative CWSS binding site of Cd.SrtB has been identified within the β6-H5 loop and contains several polar residues which could form contacts with the P1 serine of the SPKTG sorting signal together with a semi-conserved hydrophobic residue which could form contact with the P2 proline. Analysis of this region was unfortunately limited by poor electron density, presumably due to high mobility of the loop as indicated by the high B-factors of the Cα backbone atoms (which nevertheless exhibited continuous density even where side chain density was lacking). In the absence of a substrate complex structure, it is clearly not possible to determine whether β6-H5 undergoes a transition to a more ordered state upon substrate binding in a manner similar to formation of the 3_{10} helix in Sa.SrtA β6-β7. However, it is notable that in Cd.SrtB

C226A, higher than structure average B factors are observed in this region, indicating that in its *apo* form, this loop is relatively disordered.

The β7-β8 loop, which forms the putative binding site of the transpeptidation substrate (*meso*-DAPA in the case of Cd.SrtB) is of a similar overall shape to that observed in the SrtB enzymes of other species, while being slightly less extended than that of Sa.SrtB. Differences in this region are more likely to reflect differences in function rather than bacterial physiology, since the *S. aureus* SrtA enzyme has a significantly more extended β7-β8 loop than Sa.SrtB despite sharing the same polyglycine transpeptidation substrate. Given the role of Sp.SrtB in pilus biogenesis, it is therefore hardly surprising that its β7-β8 loop is radically different in shape to Sp.SrtA – its transpeptidation substrate, a pilin rather than a peptidoglycan cross-link, is likely to be more bulky and require more clearance. It is unfortunate that the *B. anthracis* structure did not exhibit electron density in this area since this species has a similar cell wall composition to *C. difficile*, having mDAP cross links; comparison between the species may have revealed specific features potentially involved in recognition of mDAP as distinct from *S. aureus* SrtB which is specific for polyglycine cross links.

In agreement with cleavage assay results indicating a lack of metal dependence presented in Chapter 8, the crystal structure of SrtB C226A did not include any bound metal ions. Furthermore, examination of the Cd.SrtB β3-β4 loop, equivalent to the Sa.SrtA metal binding site, revealed an extensive network of hydrogen bonding in place of metal co-ordination. Since stabilisation of the region is clearly achievable in the absence of metal binding, it seems likely that metal binding by *S. aureus* SrtA is a regulatory adaptation rather than a biological necessity, although metal binding may afford more robust stabilisation.

6.5.2 Failure to Crystallise Wild Type *C. difficile* SrtB

Extensive but ultimately unsuccessful efforts were made to crystallise wild type *C. difficile* SrtB. As confirmed by mass spectroscopy data, the only difference between SrtB and SrtB C226A is the intended single substitution from cysteine to alanine at position 226. Prior to solution of the SrtB C226A structure, efforts to improve crystallisation of SrtB therefore focussed on biochemical and biophysical changes which could be accounted for by the substitution.

In the low pH conditions in which SrtB C226A crystallised it is likely that the cysteine residue (with its side chain pKa of ~8.0) would be effectively entirely protonated, so charge contributions from the cysteine side chain are unlikely to interfere with crystallisation. It is quite possible, however, that Cys226 engages in thiol exchange reactions to form disulphide complexes. As was discussed in Section 4.1.1, size exclusion chromatography of SrtB indicates that SrtB is monomeric when in a buffer with neutral pH and moderate ionic strength, but conversely, results from SDS-

PAGE analysis demonstrate that that when placed in appropriate conditions SrtB is indeed capable of forming disulphide dimers, presumably due to exposure of the active site sulphydryl Cys 226 by denaturing. Crystallisation screens feature a diverse set of conditions, and it is quite possible that some of these may promote exposure of the cysteine residue and allow disulphide mediated dimers to form.

The usual method of preventing disulphide formation is by addition of reducing agents. Production of reducing conditions in crystallisation experiments is challenging due to the instability of dithiothreitol (DTT) and the volatility of 2-mercaptoethanol. In the case of Cd.SrtB it was achieved by addition of the reducing reagent TCEP. Irreversible inactivation of the cysteine residue by adduction by the known sortase inhibitor MTSET was also performed. Neither of these modifications improved crystallisation.

Failure to crystallise proteins is a common problem in X-ray crystallography, and there are many approaches to overcoming such an obstacle. Most commonly, a suitable condition can eventually be identified by continued and optimised high-throughput screening. Improving protein purity is also advised – a survey of 229 crystallisation experiments (Geerlof *et al.*, 2006) showed that proteins of >95% purity crystallise in 59% of cases wheras proteins of <95% purity only yielded 37% of cases . However, purity is unlikely to be the limiting factor in the case of Cd.SrtB as its purity is not only very high, it is essentially identical to that of SrtB C226A which crystallises readily.

Once the structure of Cd.SrtB C226A was available, it was clear that there was additional structural evidence which may help explain the failure to crystallise wild type Cd.SrtB. Specifically, the loops surrounding the active site are a key interface within crystallographic dimers of SrtB C226A. It is plausible that in the wild type enzyme, the additional bulk of the cysteine residue prevents these interactions and hence prevents dimerisation which is presumably essential for crystallisation under these conditions. Alternatively, the cysteine may prevent dimerisation by exerting effects on the general electrostatic environment of the active site.

In light of the failure of exhaustive efforts to crystallise wild type Cd.SrtB and informed by the structure Cd.SrtB C226A, it would seem worthwhile to investigate other variants of SrtB. By abolishing the contacts which are observed in SrtB C226A, but may be sub-optimal in SrtB, other contact surfaces may be favoured, some of which may be more conducive to crystallisation. Abolition of the interactions around the active site could be achieved very easily by N-terminally truncating the protein by removing residues 39 to 44 to eliminate the unstructured region which precedes H1. A more advanced method to alter the other interface between the SrtB dimers would

be to engineer the loops which form the other surface of dimerisation, β7-β8 and β6-H5. In particular, residues which are listed in Table 5.4.2 as participating in hydrogen bonding via their side chains. As a structure is available for Cd.SrtB C226A, PISA could be utilised to test loop mutants *in silico* for their potential to dimerise. Assessment of correct folding of such mutants would also be necessary, eg. by circular dichroism (CD) spectroscopy, a technique used to analyse secondary structure of proteins by the manner in which they interact with polarised light.

6.5.3 Limitations and Further Work

As with any X-ray structure determined from diffraction of only moderate resolution, vast improvements to the model could be by made by analysis of higher resolution data. Improving resolution is often achieved by further optimising crystallisation conditions or seeking new conditions which yield better performing crystals. Alternatively, a fragment of the protein may be crystallised in place of the whole protein, removing domains which may be disordered while retaining the areas of most biological significance. In addition, there are several methods for improving the diffraction of existing crystals, including annealing and removal of solvent from the crystal by dehydration (Heras and Martin, 2005). Both of these techniques aim to reduce disorder in the crystal and often reduce mosaicity as well as increasing resolution.

However, the major shortcoming of this work is not the quality of the structure but the reliance on a modelled cysteine residue for analysis of active site co-ordination. The side chain of the substituted alanine gives a good indication of main chain geometry of the cysteine, but the side chain dihedral angle was modelled by comparison with Sa.SrtB and Ba.SrtB. Serine is often chosen as a substitute for cysteine as it is viewed as the closest structural analog of cysteine, but it would be an inadvisable substitution in the case of a sortase enzyme, as a Cys to Ser substitution results in residual activity in SrtA (Frankel *et al.*, 2007b). Clearly, the solution to this problem is to crystallise the wild type SrtB protein, as previously discussed in Section 6.5.2

Another worthwhile experiment (which would not depend upon successful crystallisation of wild type Cd.SrtB) would be determine how Cd.SrtB interacts with its CWSS substrate and confirm the putative binding region in β6-H5. Introduction of ligands is frequently performed for X-ray crystallography, either before or after crystallisation (Danley, 2006) . The former requires that the ligand does not inhibit crystallisation and that the enzyme has sufficient affinity for the substrate that occupancy of the active site is high enough for the ligand to be visible in the electron density. There are no literature of a K_d for any sortase and its CWSS peptide substrate, but given that K_M of SrtA enzymes for such peptides tend to be in the millimolar range, it is likely that their affinity is poor. An additional concern is the accessibility of the active site, which is a particular concern for Cd.SrtB due to its dimerisation interfaces in the region of the active site. If a non covalent complex

is to be formed, the latter approach of post-crystallisation is therefore more likely to be successful because very high concentrations of peptide can be achieved without inhibiting crystallisation. In the case of Cd.SrtB, the use of the C226A mutant would be preferred over the wild type due to the risk of hydrolysis of the peptide by the wild type enzyme. A peptide was obtained for this purpose, but no sufficiently large SrtB C226A crystals were obtained subsequent to the crystals for which diffraction is described in this chapter.

Another approach to obtain an enzyme-substrate complex is to modify the wild type enzyme covalently. In the case of the SrtA-CWSS structure (Suree *et al.*, 2009) this was achieved by the use of a LPAT* peptide, where T* is a threonine modified with a CH_2-SH moeity in place of its carbonyl group. Disulphide formation between the T* residue and the active site cysteine results in a modification which mimics the acyl intermediate and is non-hydrolysable. Heterogeneity due to non-specific adduction would not be a concern, as Cd.SrtB contains no cysteines other than the active site residue. Such a method is likely to result in a robust level of incorporation which could be evaluated and quantified by mass spectroscopy. However, this is neither a true enzyme-substrate complex or an acyl intermediate. Such structures may be misleading, because it is possible that the initial enzyme-substrate interactions differ from those subsequent to formation of the acyl intermediate – they may be considerably weaker, since the substrate is now covalently attached to the enzyme there is less of a requirement to bind it with high affinity, while flexibility within the active site may be of benefit to allow the reaction to progress.

Structures of many enzymes have been obtained in acyl intermediate states, although this is a challenging task because such states are often inherently unstable. In cases where X-ray crystallography is to be employed, success is often possible if crystallisation and data collection can be performed rapidly, as has been possible in the case of trypsin (Mangel *et al.*, 1990) . Another approach is to halt the reaction by using a mutant enzyme incapable of resolving the intermediate state, as in the case of the *Mycobacterium tuberculosis* β-lactamase (Tremblay *et al.*, 2010). The presence of a hydrolytic shunt in the sortase mechanism rules out the latter approach, as interfering with the transpeptidation reaction will not necessarily favour accumulation of the intermediate. NMR spectroscopy is perhaps more suited to for examination of intermediates because it is possible to conduct whole experiments in the short timescales during which the intermediate is stable.

There is some evidence, presented later in this work, that quantities of acyl intermediate Cd.SrtB accumulate during reaction with an SPKTG peptide, presumably due to inefficiency of hydrolysis or transpeptidation reactions and indicating that the Cd.SrtB acyl intermediate may be relatively stable. Repurification of this product may represent a method to obtain a genuine intermediate for

crystallisation and provision of a sortase acyl intermediate structure. Other sortases may exhibit even more stable intermediates which again may be useful for determination of acyl intermediate structures.

A wider appraisal of sortase structures reveals significant gaps in the structural knowledge of sortase enzymes, perhaps not surprising for a class of enzyme discovered only 14 years ago. There is only one convincing structure of a Srt-CWSS complex structure, of a class A enzyme, and one structure of a sortase transpepdiation substrate complex, of a class B enzyme. Ideally, three structures would be determined for each class – an *apo* form, a complex with a CWSS peptide and a double complex with a CWSS and transpeptidation acceptor molecule. A single complex with the transpeptidation acceptor is of limited interest as the ping pong reaction mechanism does not favour such a complex *in vivo*.

Structures of sortases complexed with CWSS peptides may provide more information on enzyme-substrate interactions, but because the CWSS peptides which are commonly utilised only represent a very small portion of the target protein, questions remain regarding whether the enzyme interacts more widely with the target protein. Similarly with regard to the transpeptidation substrate, it is unclear whether enzyme interacts more widely with the peptidoglycan? As discussed in Section 8.4.3, the poor performance of sortase enzymes *in vitro* suggests that the short CWSS-mimic peptides are poor mimics of the *in vivo* substrates. It may therefore of benefit to explore larger fragments of CWSS-bearing proteins, to determine whether the enzyme engages in any other specific interactions with its *in vivo* substrates.

With respect to the transpeptidation substrate, a more ambitious but worthwhile experiment would be to obtain the structure of a complex with a larger peptidoglycan fragment. Such structures are available for other classes of enzyme – for example, several structures are available of complexes of bacterial peptidoglycan carboxypeptidases with their cognate peptidoglycan tetrapeptide stem (Chen *et al.*, 2009; McDonough *et al.*, 2002). In the case of *C. difficile*, synthesis of peptidoglycan fragments could prove prohibitively expensive due to the presence of *meso*-DAPA, but purification from a natural source may be feasible.

6.6 - Conclusions

An active site cysteine mutant of the Sortase B enzyme of *C. difficile* has been crystallised and its structure solved to a resolution of 2.55Å. This novel structure represents the fourth class B sortase structure to be solved, and the first structure of any sortase from the genus *Clostridium*. It exhibits a similar fold to other class B sortases, being composed of a core β-barrel flanked by several helices. Two putative substrate binding regions, for the CWSS peptide and the transpeptidation substrate, have been identified and analysed.

Plans for potential future work to gain more information from the SrtB structure have been outlined, including crystallisation of enzyme-substrate complexes and reaction intermediates. In light of the failure to crystallise a wild type *C. difficile* SrtB, a structurally-informed plan has been developed to design SrtB mutants for enhanced crystallisation.

Chapter 7 – Crystallisation and Test Diffraction of CD0386

7.1 Introduction

C. difficile CD0386 is a 111kDa protein which, as described in Chapter 5, is anchored to the *C. difficile* peptidoglycan by the sortase SrtB. Its annotation in the genome is 'Putative collagen-binding surface protein', an automatic annotation based upon sequence homology with domain B of the *S. aureus* collagen binding protein Cna. However, the N-terminal domain of CD0386 shows no sequence homology to any characterised proteins, and structural studies were planned to aid in ascribing its function.

Due to the large size of the protein, a fragment-based approach was determined most appropriate, as crystallising such large proteins can prove challenging. Secondary structure prediction and homology modelling were performed to provide speculative information on the structure of CD0386, which informed the design of a fragment of CD0386 for structural studies. This 61kDa fragment, CD0386N, was crystallised and test diffraction performed.

7.2 Methods

7.2.1 Secondary Structure Prediction and Modelling

The amino acid sequence of CD0386 was submitted to the Phyre2 Protein Fold Recognition server hosted by Imperial College, London (Kelley and Sternberg, 2009).

7.2.2 Crystallisation and test diffraction of CD0386N

CD0386N was purified as described in Section 4.3.6. Initial crystallisation conditions for CD0386N were determined by high throughput screening using a Phoenix robotic crystallisation system (Art Robbins Industries). Reservoir volumes of 50µl were dispensed into Intelliplate crystallisation plates (Art Robbins Industries) and sitting drops of 200nl assembled by dispensing protein and reservoir solution in ratios of 1:2, 1:1 and 2:1. Plates were sealed, incubated at 16°C and drops inspected microscopically after one week, then monthly. Seven commercially available screens were tested - Structure Screen I and II HT96, Clear Strategy Screen I HT96, Clear Strategy Screen II HT96, PACT Premier HT96, JCSGplus HT96, Morpheus HT96 and Heavy and Light HT96 (all obtained from Molecular Dimensions) .

Manual optimisation of the condition 0.2M KSCN, 0.1M sodium cacodylate pH6.5 and 15% PEG 4k was performed by the hanging drop method in a 16 well tissue culture plate with a reservoir

volume of 1mL, sealed with greased 20mm siliconised glass cover slips. Variables optimised were pH of crystallisation solution (pH 5.5-7.0), and KSCN concentration (175-300mM).

Test diffraction was performed at beamline I-04 of the Diamond Light Source and diffraction recorded with an ADSC Q315 CCD detector.

7.3 Results

7.3.1 Secondary Structure Predication and Phyre2 modelling of CD0386

Domain analysis by Phyre2 is shown in Figure 7.3.1 and Table 7.3.1

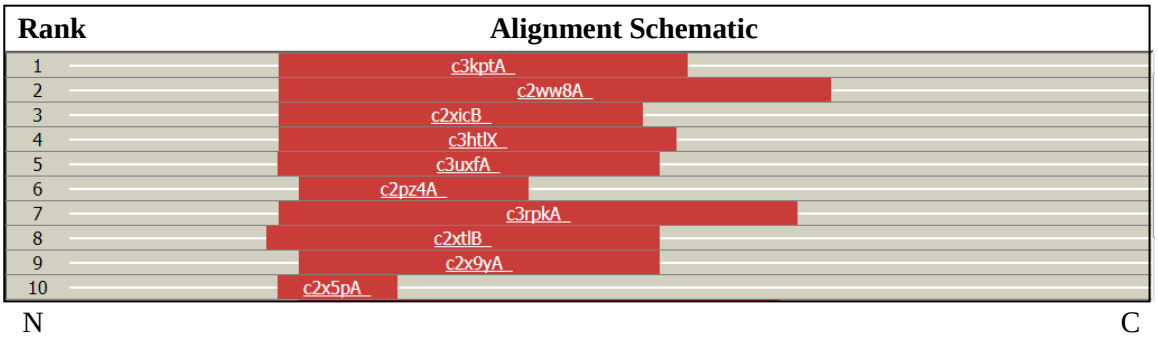
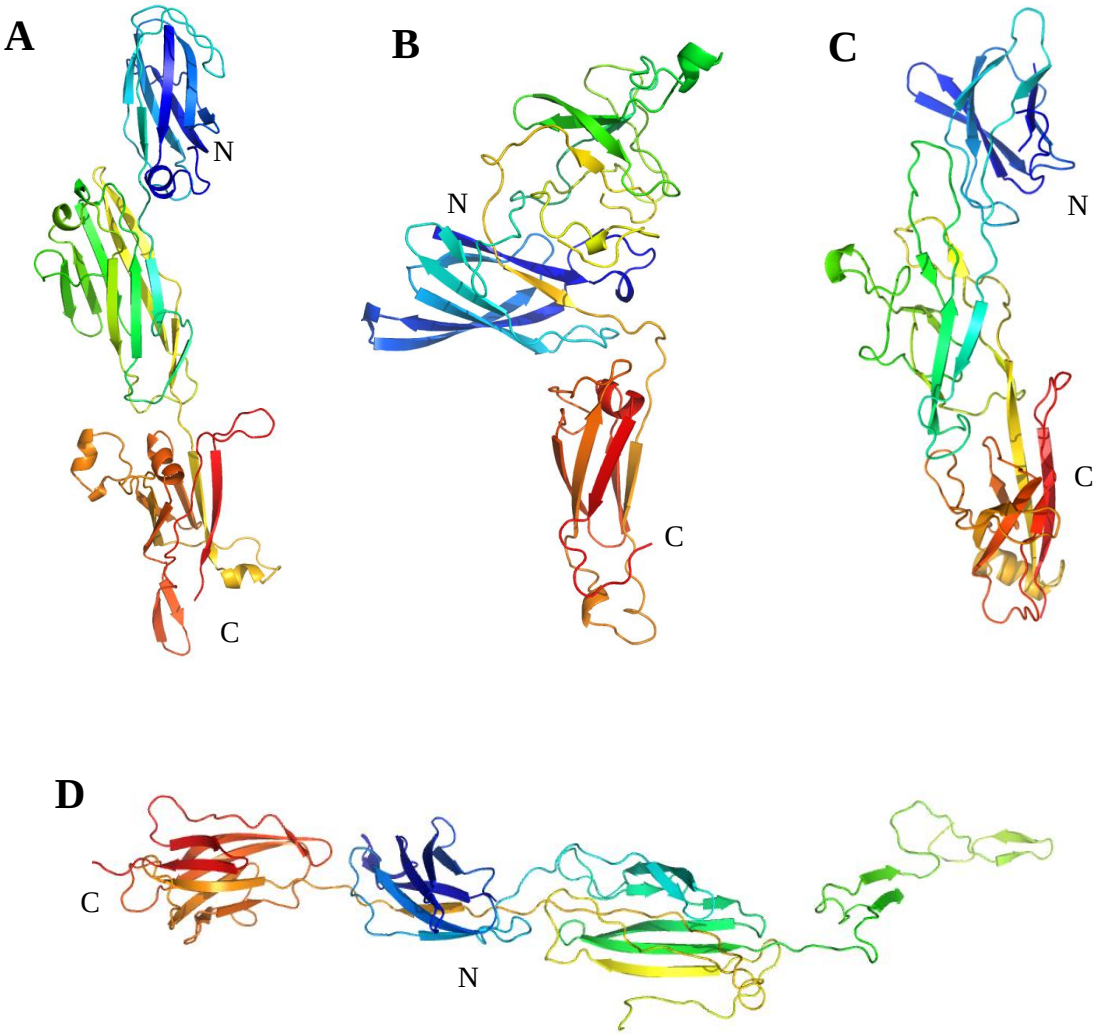


Figure 7.3.1 – Phyre2 Analysis of CD0386 - Alignment of the 10 highest scoring profile-profile matches

Rank	PDB Code	Species	Description	Reference
1	3KPT	*B .cereus*	BcpA, major pilin	(Budzik *et al.*, 2009)
2	2WW8	*S. pneumoniae*	RrgA, pilus adhesin	(Izoré *et al.*, 2010)
3	2XIC	*S. pyogenes*	SPY0125, pilus adhesin	(Pointon *et al.*, 2010)
4	3HTL	*C. diptheriae*	SpaA, major pilin	(Kang *et al.*, 2009)
5	3UXF	*A. oris*	FimP, Fimbral pilin	(Persson *et al.*, 2012)
6	2PZ4	*S. agalactiae*	GBS52, minor pilin, adhesin	(Krishnan *et al.*, 2007)
7	3RPK	*S. pneumoniae*	RrgB, major pilin	(Paterson and Baker, 2011)
8	2XTL	*S. agalactiae*	BP-2a, major pilin	(Necchi *et al.*, 2011)
9	2X9Y	*S. pneumoniae*	RrgB, major pilin	(Spraggon *et al.*, 2010)
10	2X5P	S. pyogenes	Fbab-B	(Oke *et al.*, 2010)

Table 7.3.1 – Phyre2 Analysis of CD0386 - Details of the 10 highest scoring profile-profile matches

Models of CD0386 produced by Phyre2 based upon the four highest-scoring profile-profile matches are shown in Figure 7.3.2.



7.3.2 – *In silico* Modelling of CD0386. Homology/threading models of CD0386 produced by Phyre2 based upon (A) *B. cereus* BcpA (B) *S. pyogenes* SPY0125 (C) C. *diptheriiae* SpaA (D) *S. pneumoniae* RrgA

7.3.2 – Design of CD0386N

CD0386N was designed to encompass the area with predicted structural homology to bacterial pilins and the N-terminal region with no detected structural homologues. Its amino acid sequence is detailed in Appendix A. Cloning, expression and purification of CD0386N was as described in Chapter 4 (Section 4.3.6).

7.3.3 Crystallisation and test diffraction of CD0386N

Crystals were observed in well C11 of the Clear Strategy Screen 1, a condition composed of 0.2M KSCN, 0.1M sodium cacodylate pH6.5 and 15% PEG 4k (Figure 7.3.3 A). Optimisation to pH 5.5 yielded larger clusters of crystals (Figure 7.3.3 B) which were broken and subject to X-ray diffraction at Beamline I-04 of the Diamond Light Source on 30th May 2012. (Figure 7.3.3 C).

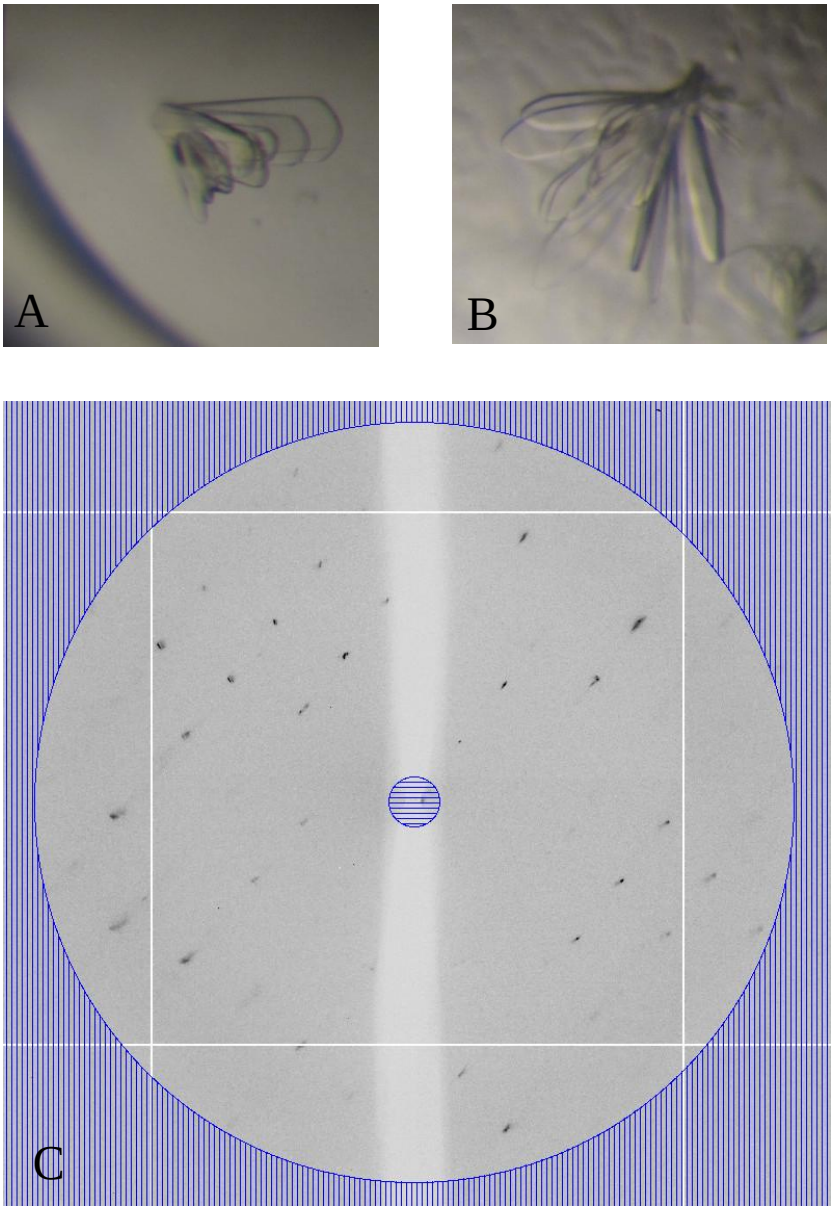


Figure 7.3.3 – Crystallisation and Test Diffraction of CD0386N. (A) Crystals grown in 0.2M KSCN, 0.1M sodium cacodylate pH6.5 and 15% PEG 4k. (B) 0.2M KSCN, 0.1M sodium cacodylate pH5.5 and 15% PEG 4k. (C) Test diffraction of crystals shown in panel B (the hatched circle represents a resolution of 8Å).

7.4 Analysis and Discussion

7.4.1 Secondary Structure Predication and *in silico* Modelling of CD0386

The Phyre2 program performs prediction of protein structures based upon secondary structure prediction, homology modelling and threading. At the core of the system is a 'fold library' constructed by analysis of solved structures deposited in the Structural Classifications of Proteins (SCOP) database (Murzin *et al.*, 1995). Upon submission of a protein sequence to Phyre2, the server uses three separate programs to predict the secondary structure of the query protein. This secondary structure profile is combined with a PSI-BLAST search of the fold library to identify distant sequence homologues. The combined secondary structure and PSI-BLAST profile of the query protein is then used to search the fold library by profile-profile matching, identifying and scoring proteins with predicted structural homology. Homology modelling and threading is then performed including modelling of insertions and deletions.

Phyre2 identified with high confidence several structural homologues of CD0386N. All of the highest scoring matches are with structures of bacterial pilins, either major structural pilins or minor functional pilins. However, it is noteworthy that CD0386 does not appear to be pilus-associated in *C. difficile*, as it is monomeric when isolated from cell walls (see Section 5.3.5) and does not contain the characteristic 'YPKN' motif which acts as transpeptidation acceptor nucleophile in the pilus-associated sortase reaction (Budzik *et al.*, 2008a)

Two of the structures identified as putative homologues of CD0386 exhibit unusual structural features – the BcpA pilin of *B.cereus* contains two intramolecular amide ('isopeptide') bonds, while the *S. pyogenes* contains two isopeptide bonds and an intramolecular thioester bond. These features are presumably required for stabilisation of the proteins to fulfil their functions. The formation of amide bonds results in a loss of mass of 18Da corresponding to elimination of a water molecule – such a mass change is not observed in CD0386 (see Section 4.3.6) but may be masked by other modifications.

Modelling of CD0386 by Phyre2 produces structures which appear superficially similar, each featuring an overall linear structure composed of two or three β-fold domains. However, closer inspection reveals significant differences in the domain arrangement, with the linear structure formed by either a contiguous (eg. Fig. 7.3.2 A) or non-contiguous (eg. Fig. 7.3.2 C) arrangement of the β fold domains. This highlights a particular shortcoming of *in silico* modelling, as the program has confidently identified the folds of discrete domains, but is not able to predict how these domains are arranged to form the whole protein.

Based upon this i*n silico* analysis, a 61kDa fragment was designed (CD0386N) encompassing both the pilin-like domain and the N-terminal domain for which Phyre2 did not detect any fold homology.

7.4.2 Crystallisation and Test Diffraction of CD0386

CD0386N crystallised readily in the conditions of 0.2M KSCN, 0.1M sodium cacodylate pH5.5 and 15% PEG 4k. X-ray diffraction was performed, but diffraction was poor, with a low number of reflections and with a maximum resolution of <8Å. However, the observed diffraction pattern was characteristic of protein diffraction.

Having established that the crystals grown under these conditions are proteinaceous, further optimisation is required - general discussion on the optimisation of crystallisation was provided previously, in Section 6.5.2.

7.5 Conclusions

Preliminary *in silico* analysis of the sortase-anchored protein CD0386 suggests that it possesses homology to several bacterial pilins, several of which perform adhesin functions. A 61kDa N-terminal fragment of CD0386 (named CD0386N) has been crystallised and subjected to test diffraction. Further optimisation of crystallisation is very likely to lead to solution of a structure of CD0386, providing insight into the potential role of this protein in pathogenesis.

Chapter 8 – Biochemical Characterisation of the *C. difficile* SrtB reaction

8.1 Introduction

8.1.1 Sortase Assay Methodologies

Initial investigations of sortase activity (eg. Ton-That *et al.*, 1999) employed a detection technique based on Förster Resonance Energy Transfer (FRET). The principle of FRET is based upon the observation that if a fluorescent molecule emits light at a wavelength overlapping the excitation wavelength of a sufficiently proximal acceptor fluorophore, the acceptor molecule will be stimulated to fluoresce (Gadella, 2009). The transfer of energy is via a 'virtual photon' rather than emitted light, hence why the commonly-used description of '*Fluorescence* Resonance Energy Transfer' is considered potentially misleading.

As the efficiency of energy transfer between the donor and acceptor pair depends upon the distance between them, FRET can be used as a dynamic 'molecular ruler'. Indeed, the first proof-of principle implementation of FRET in a biological context was to measure the lengths of polyprolyl peptides ranging from 12Å to 46Å in length, confirming that FRET efficiency is a function of the inverse sixth power of the distance between the donor and acceptor (Stryer and Haugland, 1967). In its simplest implementations, FRET can be used as a purely qualitative indicator of whether two molecules are in close proximity. This is particularly useful for peptide cleavage assays – the length of a typical peptidase probe is typically short enough for FRET to occur between fluorophores coupled to the N- and C-termini of the peptide. Cleavage of the probe separates the two fluorophores, and as they diffuse in the reaction solvent, FRET is lost and fluorescence of the acceptor is lost. With the availability of fluorimetric microtitre plate readers, the approach is highly amenable to high throughput methodologies.

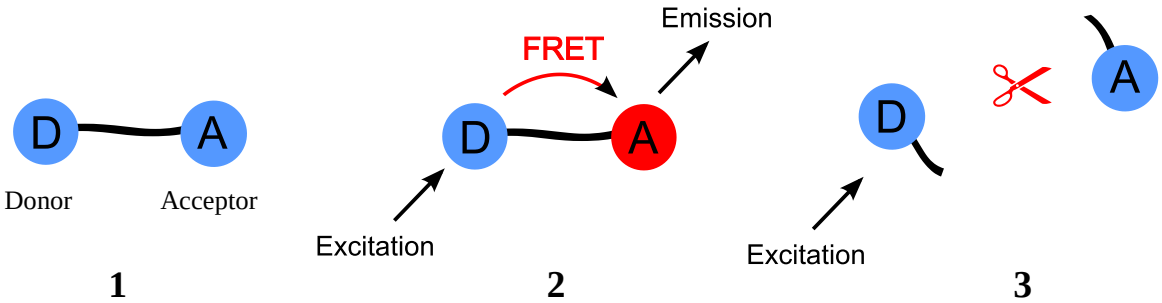


Figure 8.1 – FRET as a technique for assaying cleavage of a peptide probe. (1) Donor and acceptor are held in close proximity by the probe molecule. (2) Upon excitiation of the donor, FRET results in fluorescence of the acceptor molecule. (3) Cleavage of the probe results in a loss of FRET coupling and hence a reduction in fluorescence of the acceptor.

There is, however, an inherent weakness in the use of FRET for the study of enzymes. At the very high substrate concentrations which are often necessary in K_M determination experiments, probe molecules engage in non-specific absorption, whereby the donor moiety of one peptide engages in FRET with the acceptor of another discrete molecule. These 'inner filter effects' severely limit the utility of FRET for quantitative kinetic studies but it remains a very useful tool for qualitative studies, eg. probing substrate specificity and for inhibitor studies where a high throughput methodology is required. An additional limitation of FRET in the context of sortase enzymes is that it only assays *cleavage* of the CWSS-mimic probe – formation of the transpeptidation product cannot be assessed.

Early kinetic characterisation of *S. aureus* SrtA activity was performed by FRET in spite of the above limitations. Using a probe mimicking the 'LPETG' SrtA CWSS coupled to Dabcyl and Edans, K_M of *S. aureus* for LPETG was determined to be 16.5 µM and k_{cat} was determined to be $2.3x10^{-5}$ s^{-1} (Ton-That *et al.*, 2000). However, when this work was repeated using a High Performance Liquid Chromatography (HPLC) method to robustly quantify the reaction products, it was found that the parameters reported by the FRET experiments were substantially erroneous, confirmed by the observation of quenching effects at fluorescent substrate concentrations above 25µM. The revised K_M was 5.5mM while the true k_{cat} was determined to be 0.27 s^{-1} (Kruger *et al.*, 2004c). Both parameters represent remarkable increases from the initially reported figures.

8.1.2 - Introduction to Liquid Chromatography Coupled Mass Spectroscopy (LC-MS)

Liquid chromatography coupled mass spectroscopy (LC-MS) has rapidly become a standard technique in analytical laboratories and is now finding wide and varied applications in research. Mass spectroscopy is not well suited to analysing complex samples, but by separating samples via HPLC directly prior to analysis, this weakness is overcome by the extraordinary ability of HPLC to separate molecules. Typically, the HPLC step is performed on an ultra high performance (uHPLC) reverse phase column with a bore of <1mm. On exiting the column, the sample flow passes directly into the electrospray ionisation source of a mass spectrometer which may be of the time of flight (TOF) type for determination of mass/charge ratio or a more complex multiple-quadropole equipped instruments (LC-MS/MS) where further structural information is acquired. UV detection between the column and MS instrument is very common, providing even more information on the analytes. The limitations of LC-MS are essentially those of its constituent technologies – HPLC is unable to separate highly similar molecules such as enanantiomers, while MS is unable to discriminate isobaric molecues or to analyse compounds which cannot be efficiently ionised.

8.1.3 Aims of the work Described in this Chapter

This chapter describes characterisation of the activity of *C. difficile* SrtB *in vitro*. The aims were to:

1. Confirm that *C. difficile* SrtB cleaves an SPKTG signal as suggested by Comfort and Clubb (2004) and to confirm or exclude the cleavage of an NVQTG signal as speculatively proposed by Tulli *et al* (2013).
2. Demonstrate that the enzyme is a functional sortase, capable of the whole transpeptidation reaction with meso-diaminopimelic acid (mDAP), the diamino acid which forms the cross-link of the *C. difficile* peptidoglycan.
3. Characterise the kinetics of the *C. difficile* SrtB reaction (although this was ultimately limited to provisional experiments).

8.2 - Methods

8.2.2 – Forster Resonance Energy Transfer (FRET) Assay of *C. difficile* SrtB

Probe peptides (Table 8.2.1) were synthesised by Peptide Protein Research Ltd by the solid phase Fmoc method.

Peptide Sequence	N-terminal Modifications	C-terminal Modifications	Molecular Mass
SDSPKTGSD	Dabcyl	Edans	1392.46
SDNVQTGDS	Dabcyl	Edans	1421.45
SDTKSGPDS	Dabcyl	Edans	1392.46

Table 8.2.1 – Peptides used in FRET Sortase Assays. Dabcyl = 4(dimethylaminoazo) benzene-4-carboxylic acid. Edans = 5-((2-Aminoethyl)amino) naphthalene-1-sulfonic acid

Peptides were diluted to 50µM in reaction buffer (50mM HEPES pH7.5, 250mM NaCl). Enzyme (SrtB or SrtB C226A) was added to achieve a final concentration of 5µM in a reaction volume 100µL. Reactions were assembled in the wells of a black, non-fluorescent 96-well plate and incubated at 37ºC.

Fluorescence readings were taken at 0hrs and after 16hrs incubation, using a fluorimetric plate reader (BMG Fluostar) which was configured with an excitation wavelength of 320nm and detection wavelength range of 500-540nm. The fluorescence readings from the 'buffer only' wells were subtracted from all other wells, following which the increase in fluorescence was calculated by subtracting the values of the 0hrs plate from those after 16hr of incubation.

8.2.3 Development of a HPLC Assay for *C. difficile* SrtB

A range of probe peptides (Table 8.2.2) were synthesised by Peptide Protein Research Ltd by the Fmoc solid phase method.

N-terminus	Sequence	C-terminus	Molecular Mass
Abz	SPKTG	Dap(Dnp)	860
-	G	Dap(Dnp)	327
Abz	SPKT	-	550
Abz	NVQTG	Dap(Dnp)	803
-	YSSSPKTGSSY	-	1162
-	YSSNVQTGSSY	-	1192

Table 8.2.2 – Peptides used in HPLC Sortase Assays. Where Abz = Aminobenzoic Acid, and Dap(Dnp) = Diaminoproprionic dinitrophenol.

High Performance Liquid Chromatography (HPLC) was performed using an Akta purifier system (GE Healthcare) coupled to a Supelco octadecyl silica reversed phase column of the dimensions 4.6mm x 250mm and a particle size of 3µm. Throughout chromatographic operations, flow rate was 1ml/min and absorbance was recorded at 320nm, 370nm and where indicated, 280nm. Peak areas were calculated by integration using the evaluation module of the Akta unicorn suite and with subtraction of a zero baseline. Mobile phases were composed of H_2O, 0.1% TFA (Buffer A) and ACN, 0.01% TFA (Buffer B). Two chromatographic programmes were developed:

Short Program: The column was equilibrated with 20% Buffer B, 80% Buffer A. A 100µL sample loop was flushed with Buffer A then loaded with sample in a volume of 40uL. The injection valve was actuated and after flushing of the loop with 1ml of mobile phase, a gradient from 20% to 32% Buffer B was initiated over 12 minutes. Following the gradient endpoint, the column was flushed with 1 column volume of 100% Buffer B.

Long Program: The column was equilibrated with 5% Buffer B, 95% Buffer A. A 100uL sample loop was flushed with Buffer A then loaded with sample in a volume of 40uL. The injection valve was actuated and after flushing of the loop with 1ml of mobile phase, a gradient from 5% to 40% Buffer B was initiated over 30 minutes. Following the gradient endpoint, the column was flushed with 1 column volume of 100% Buffer B.

8.2.4 Validation of Chromatographic Method for HPLC Assay of *C. difficile* SrtB

To determine a) the linearity of the relationship between Abz-SPKTG-(Dap)Dnp and peak area and b) the reproducibility of the HPLC analysis, a range of standards from 1000-100µM were prepared and analysed in triplicate by the 'short' method.

8.2.5 – HPLC Sortase Assay General Methods

Reactions were performed in a buffer of 50mM Tris pH7.5, 150mM NaCl. Peptide stocks (100mM) stored at -80°C were thawed and diluted to the appropriate concentration in chilled reaction buffer. Where mDAP was required in a reaction to mimic the cell wall transpeptidation substrate, it was diluted to the appropriate concentration in the peptide solution. The appropriate volume of SrtB stock, an equivalent mass of SrtB C226A or an equivalent volume of buffer was added to chilled reaction tubes. Where MTSET was required in the final reaction, this was added to the buffer immediately before the addition of SrtB to ensure rapid inactivation of the enzyme in spite of the instability of MTSET in aqueous conditions. Reactions were mixed by vortexing, centrifuged for 10s, then incubated at 37°C in a thermostatic water bath. Sampling was performed by removal of 20µl aliquots from the reaction tubes and quenching by rapid mixing with 25mM HCl. Samples were frozen at -80°C until analysis.

8.2.6 Generation of Progress Curves

Abz-SPKTG-Dap(Dnp) at a concentration of 250µM was incubated with SrtB at concentrations of 50µM, 100µM and 200µM. For each timepoint, three negative control samples were prepared, containing buffer only, SrtB C226A or SrtB C226A. Samples were taken at 1, 2, 3, 5, 8, 12 and 24hrs and analysed by the short method.

8.2.7 Effect of Metal Ions on SrtB Activity

A range of of aqueous salts solutions were prepared at 500mM concentration - NaCl, KCl, $CaCl_2$, MgCl, $MnCl_2$, $ZnCl_2$ and $FeCl_3$ and $(NH_4)_2SO4$, in addition to a 50mM solution of ethylenediaminetetraacetic acid (EDTA). Assays were performed in a volume of 70µL, where the final concentration of the additive salt was 50mM, Abz-SPKTG-(Dap)Dnp of 250µM and a SrtB concentration of 50µM.

8.2.8 Analysis of *C. difficile* SrtB Reactions by LC-MS

Liquid Chromatography Mass Spectrosopy (LC-MS) was performed by Dr Anneke Lubben of the Chemical Characterisation and Analysis Facility, University of Bath.

Ultra High Performance Liquid Chromatography (uHPLC) was performed with a Dionex Ultimate 3000 system coupled to an octadecyl uHPLC column of the dimensions 5mm x 2.1mm (ACE HPLC), at a flow rate of 1ml/min and a column temperature of 30°C. Mobile phases were Buffer A H2O, 0.1% formic acid (Buffer A) and ACN, 0.1% formic acid (Buffer B). Gradient elution was performed from 2-40% Buffer B over 19mins.

UV Detection of analytes was performed at 320nm for Abz-SPKTG-Dap(Dnp) and 280nm for YSSSPKTGSSY. Subsequent to UV detection, flow splitting was performed to reduce the flow rate entering the to 0.2ml/min, following which the eluate flow was analysed by ElectroSpray Ionisation Mass Spectroscopy (ESI-MS) using a Bruker Microtof Q instrument in full scan mode (+/- 50-1500 m/z). Calibration of the ESI-MS instrument with sodium formate was performed prior to analysis of the samples. Analysis of spectra was performed using Bruker Compass suite.

8.3 Results

8.3.1 – Preliminary FRET assay of *C. difficile* SrtB

Peptides Dabcyl-SPKTG-Edans, Dabcyl-TKSGP-Edans and Dabcyl-NVQTG-Edans were incubated for 16hrs in the presence of either SrtB or SrtB C226A. An increase in fluorescence relative to the SrtB C226A mutant was observed for SPKTG only. As shown, the results exhibited unacceptably high deviations from the mean. For this reason, and to avoid potential inner filter effects, the FRET assay was abandoned in order to develop a more robust, HPLC-based assay.

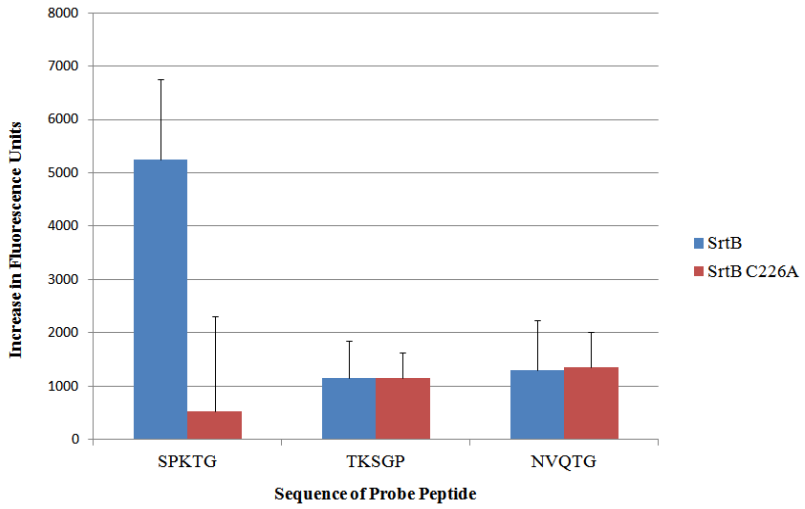


Figure 8.3.1 – Forster Resonance Energy Transfer (FRET) Assay of *C. difficile* SrtB. Mean increase in fluorescence upon incubation of self quenching peptides with SrtB or SrtB C226A.

8.3.2 Development of a HPLC Assay for C. difficile SrtB

To select appropriate wavelengths for monitoring of Abz-SPKTG-Dap(Dnp) and its cleavage products, UV spectra were obtained for Abz-SPKTG-Dap(Dnp) and its predicted cleavage products SPKTG-Dap(Dnp) and G-Dap(Dnp). On the basis of this analysis, 320nm was selected for detection of Abz-SPKT and 370nm for detection of Abz-SPKTG-Dap(Dnp) and G-Dap(Dnp).

HPLC method development was performed empirically using the method of (Kruger *et al.*, 2004b) as a starting guideline. A chromatographic method was developed to resolve all three analytes. However, the run time (including column equilibration) was ~40mins, which was deemed unacceptably lengthy considering the quantity of samples envisaged. The 'short' method was therefore developed, to capture only Abz-SPKTG-Dap(Dnp) and G-Dap(Dnp) with a run time of ~20mins.

Example separations of a mixture of standards corresponding to the expected cleavage products of Abz-SPKTG-Dap(Dnp) are shown in Figure 8.3.3 (Short method) and Figure 8.3.4 (Long method).

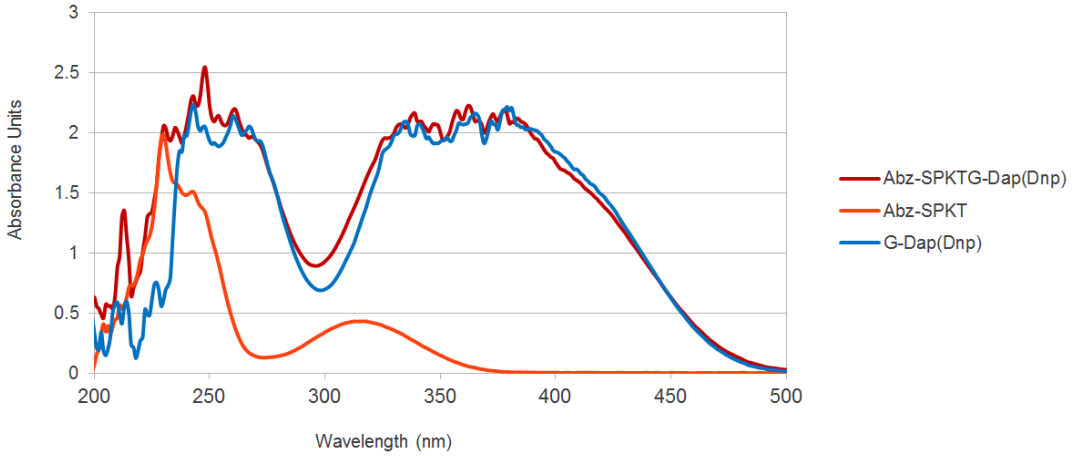


Figure 8.3.2 – UV/Visible Spectra of Abz-SPKTG-Dap(Dnp) and Cleavage Product Standards. The absorbance of a 7mM solution of each peptide was determined at wavelengths between 200 and 500nm using a nanovolume spectrophotometer.

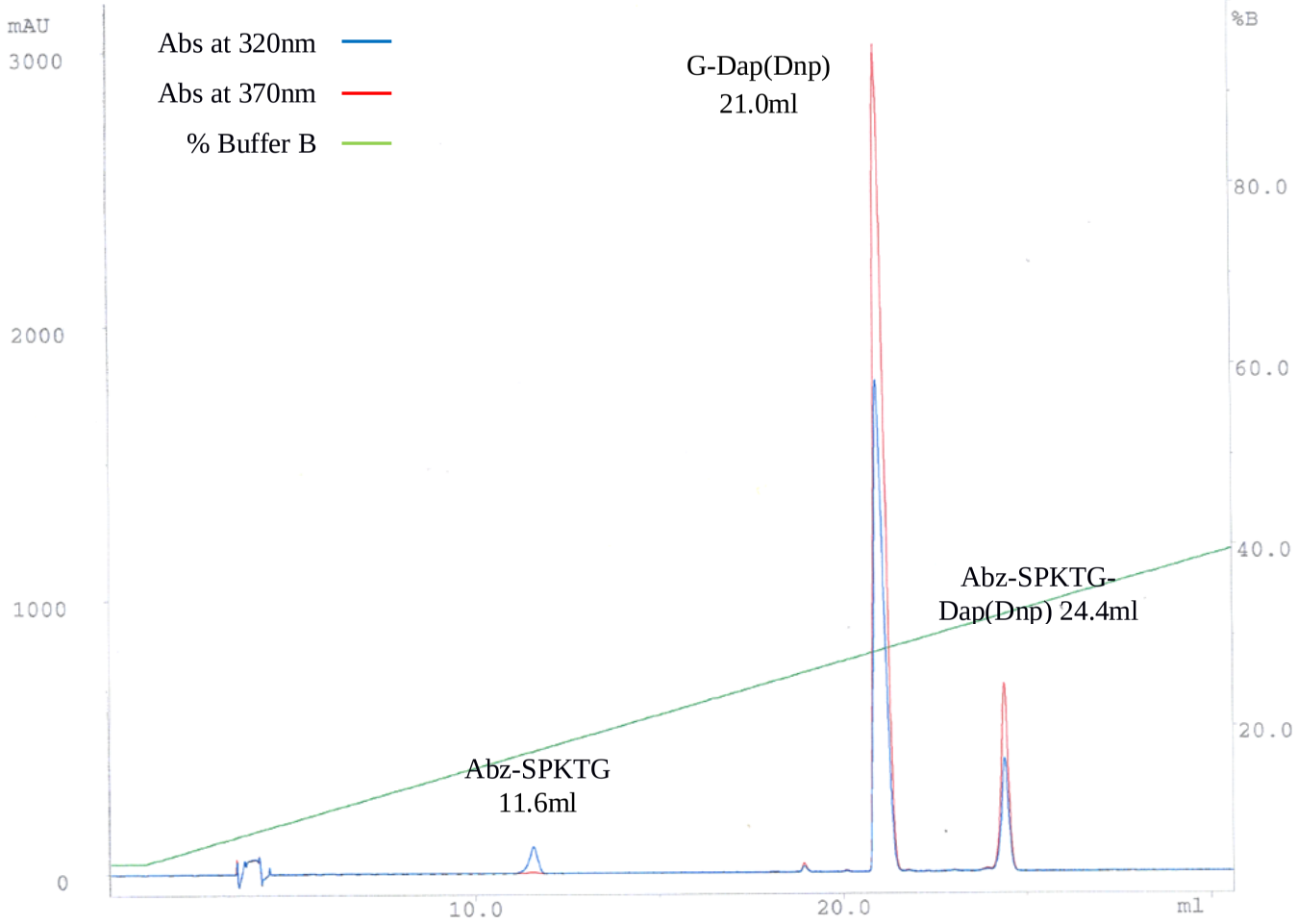


Figure 8.3.3 – Chromatography of Abz-SPKTG-Dap(Dnp) and its cleavage product standards by the Long Method. A mixture of standards composed of 10mM G-Dap(Dnp), 1mMAbz-SPKTG-Dap(Dnp) and 1mM Abz-SPKTG was analysed by the 'long' method described in Section 8.2.3. Note tailing of the G-Dap(Dnp) peak is due to the limited solubility of this compound necessitating the presence of 50% DMSO in the sample buffer.

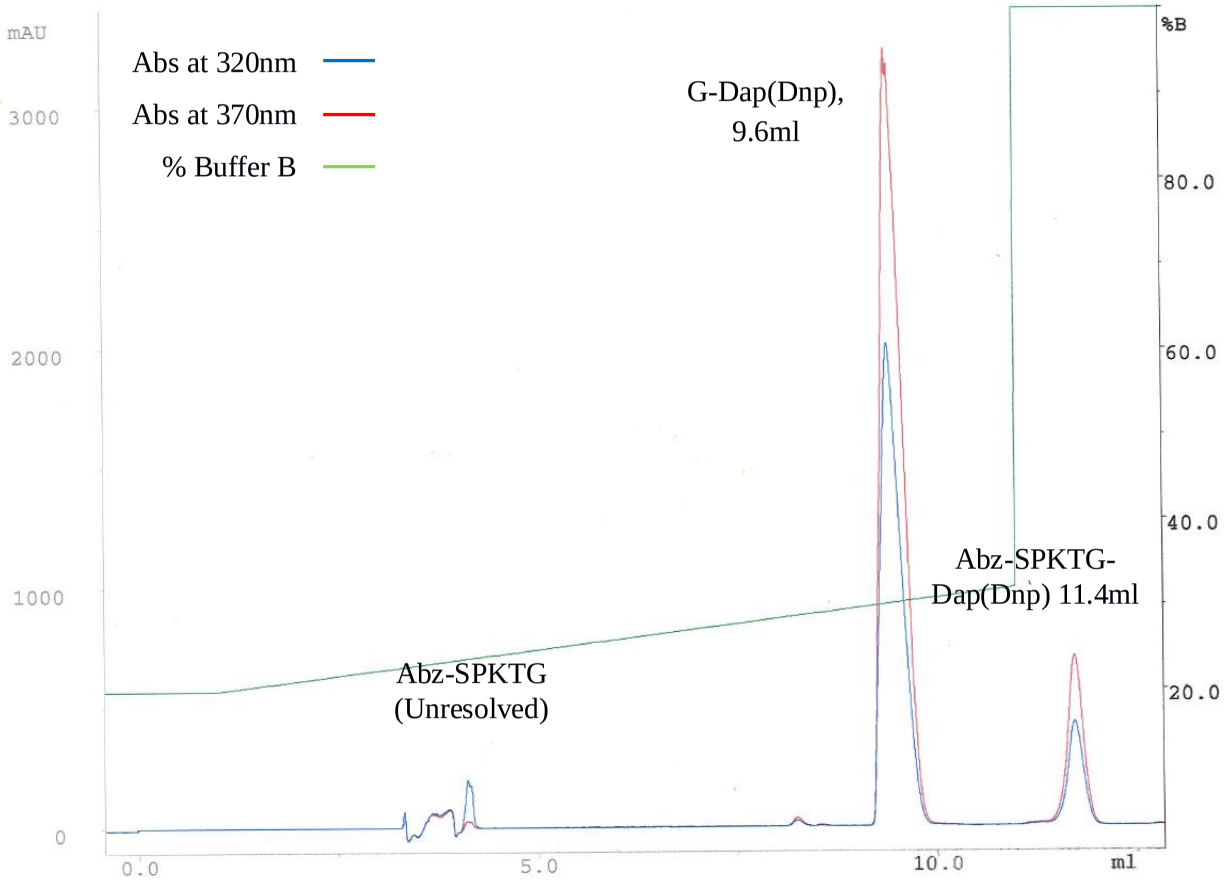


Figure 8.3.4 – Chromatography of Abz-SPKTG-Dap(Dnp) and its cleavage product standards by the Short Method. A mixture of standards composed of 10mM G-Dap(Dnp), 1mMAbz-SPKTG-Dap(Dnp) and 1mM Abz-SPKTG was analysed by the 'short' method described in 8.2.3. Note tailing of the G-Dap(Dnp) peak is due to the limited solubility of this compound necessitating the presence of 50% DMSO in the sample buffer.

8.3.3 Standard Curve for Quantification of Abz-SPKTG-Dap(Dnp)

To confirm the linearity of the relationship between Abz-SPKTG-Dap(Dnp) concentration and peak area and the reproducibility of the analysis, a range of standards from 100-1000µM were analysed in triplicate by the short method in a volume of 40µl (Figure 8.3.5).

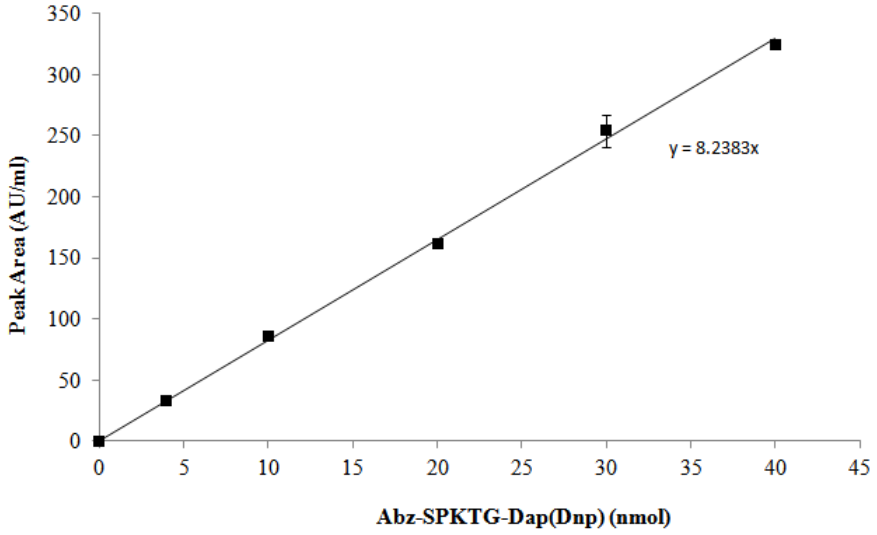


Figure 8.3.5 – Quantification of Abz-SPKTG-Dap(Dnp) by HPLC. The HPLC method facilitates reproducible, linear quantification in the range of 4nmol to 40nmol. Mean values are shown (n=3) with error bars representing +/- 1SD.

The slope of the standard curve show in Figure 8.3.5 represents the peak area per nanomole of Abz-SPKTG-Dap(Dnp). This peak area coefficient of 8.24AU.ml^{-1}.nmol^{-1} was used for all subsequent quantification. Raw data are shown in Table 8.3.1.

Concentration of Standard (mM)	Molar quantity per 40µl (nmol)	Peak Area (AU.ml^{-1})	
		Mean	Standard Deviation
0	0	0.00	0.00
0.1	4	33.17	0.97
0.25	10	86.03	4.13
0.5	20	161.63	3.08
0.75	30	254.03	13.42
1	40	325.00	0.36

Table 8.3.1 – Chromatographic Analysis of Abz-SPKTG-Dap(Dnp) standards.

8.3.4 Incubation of SPKTG and NVQTG peptides with SrtB

Abz-SPKTG-Dap(Dnp) and Abz-NVQTG-Dap(Dnp) peptides were incubated at a concentration of 250µM with either 100µM SrtB, 100µM SrtB C226A or 100µM SrtB treated with 1mM MTSET. After 16hrs incubation at 37°C, reactions were quenched with an equal volume of 25mM HCl and analysed by HPLC using the 'short' method. Molar quantities of product per reaction were calculated by multiplying the peak area corresponding to G-Dap(Dnp) by 8.24 and scaled to the whole reaction volume of 60µl by multiplying by 3. Results are shown in Figure 8.3.6 and Table 8.3.2.

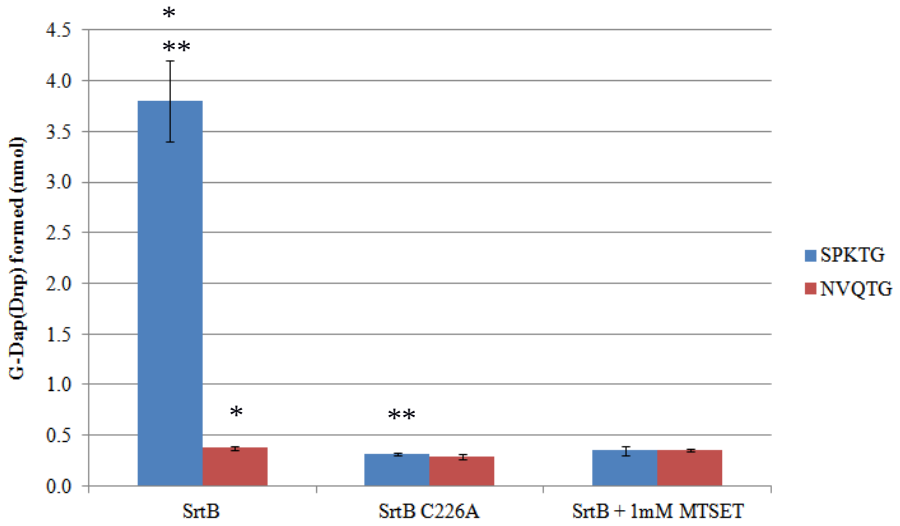


Figure 8.3.6 – Cleavage of SPKTG and NVQTG Peptides by SrtB over 16hrs. Mean values (n=3) are plotted with error bars representing +/- 1SD. Asterisks (*) and (**) indicate a p value of <0.01 as determined by Student's T-test.

Peptide Sequence	Assay Condition	Quantity of G-Dap(Dnp) Formed (nmol)				
		Rpt 1	***Rpt 2***	***Rpt 3***	***Mean***	***Stdev***
SPKTG	*SrtB*	4.296	3.786	3.313	3.799	0.401
	SrtB C226A	0.291	0.328	0.328	0.316	0.017
	SrtB + 1mM MTSET	0.364	0.400	0.291	0.352	0.045
NVQTG	*SrtB*	0.364	0.364	0.400	0.376	0.017
	SrtB C226A	0.328	0.255	0.291	0.291	0.030
	SrtB + 1mM MTSET	0.328	0.364	0.364	0.352	0.017

Table 8.3.2 – Cleavage of SPKTG and NVQTG Peptides by SrtB over 16hrs.

8.3.5 Cleavage of Abz-SPKTG-Dap(Dnp)/YSSSPKTGSSY at various concentrations of SrtB

To select an optimum enzyme concentration for subsequent experiments, Abz-SPKTG-Dap(Dnp) was incubated with SrtB at a range of concentrations from 0-200µM. Cleavage of YSSSPKTGSSY was also tested. As shown, substantial cleavage of Abz-SPKTG-Dap(Dnp) was achieved in a dose dependent manner whereas cleavage of YSSSPKTGSSY was minimal in all cases.

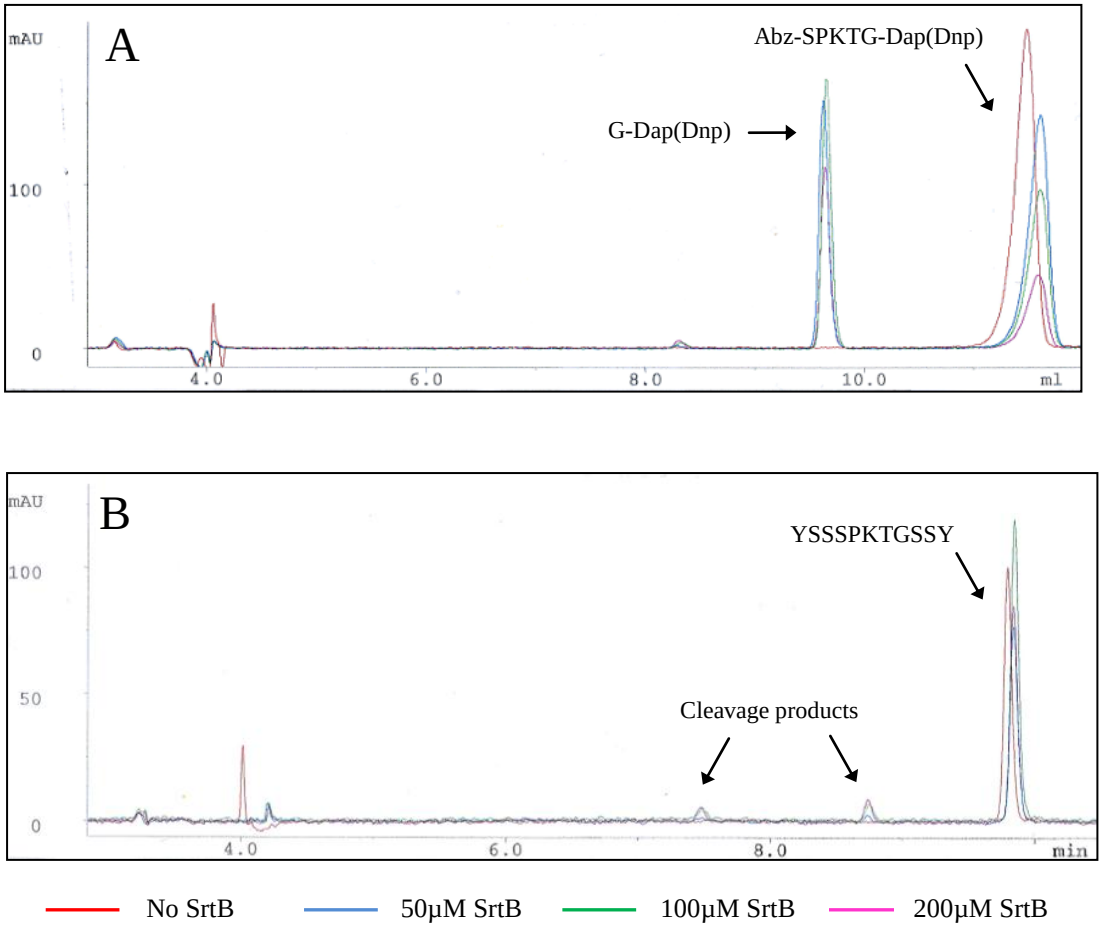


Figure 8.3.7 – Cleavage of Abz-SPKTG-Dap(Dnp) (A) and YSSSPKTGSSY (B) Peptides by SrtB. Peptides at a concentration of 250µM were incubated for 16hrs with 0, 50µM, 100µM or 200µM SrtB.

8.3.6 Preliminary Progress Curves of SrtB cleavage of Abz-SPKTG-Dap(Dnp)

To obtain provisional information on the kinetics of SrtB cleavage of the CWSS-mimetic peptide, Abz-SPKTG-Dap(Dnp) was incubated with SrtB at a concentration of 50µM, 100µM and 200µM in the presence or absence of 1mM mDAP. Due to the necessity to conserve enzyme stocks, it was only possible to collect triplicate data for the curve at the highest enzyme concentration.

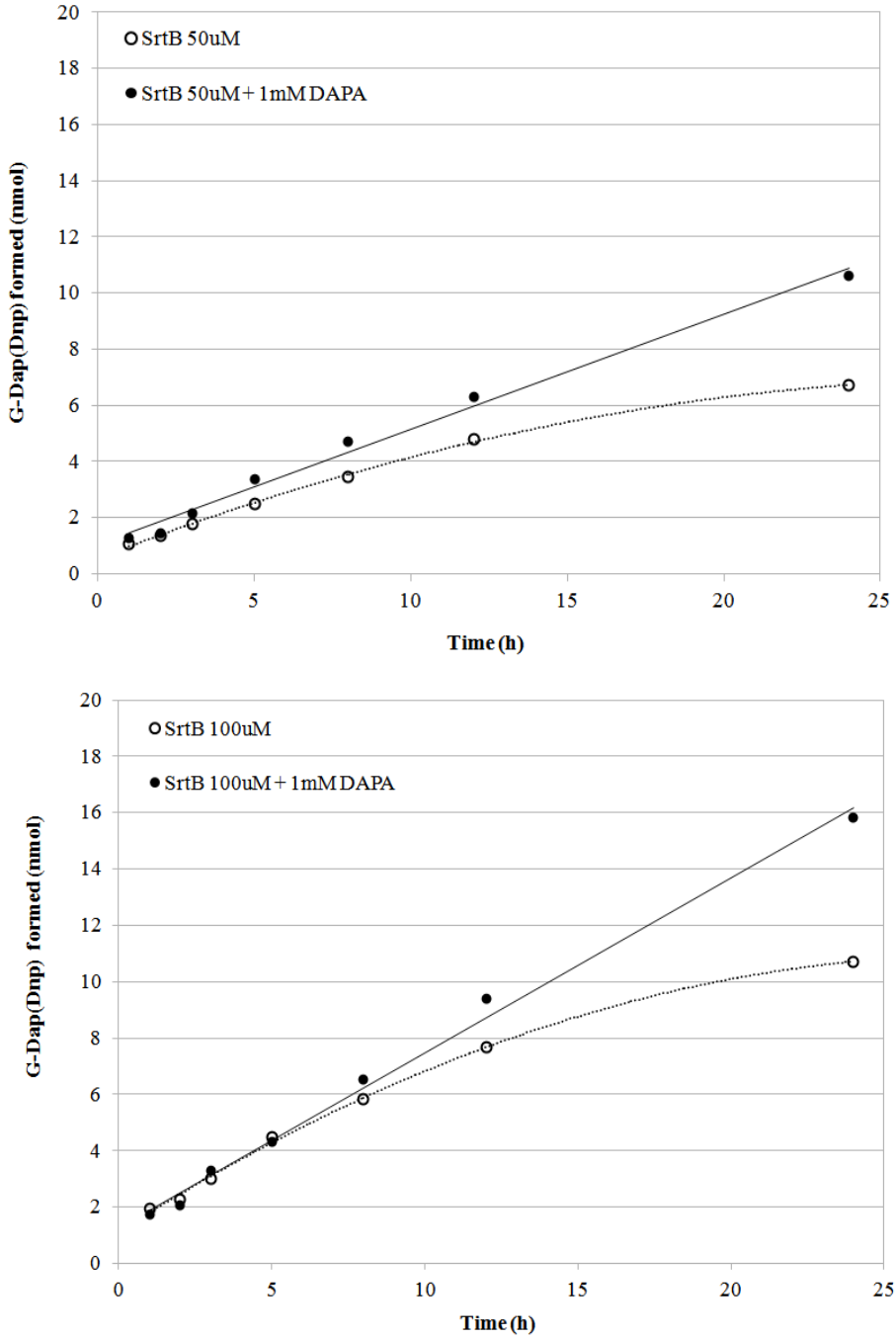


Figure 8.3.8 - Progress Curves of Cleavage of Abz-SPKTG-Dap(Dnp) by SrtB at 50µM and 100µM. The open circles (SrtB) are fitted with a second order polynominal curve, while the filled circles (SrtB + mDAP) are shown with a linear fit.

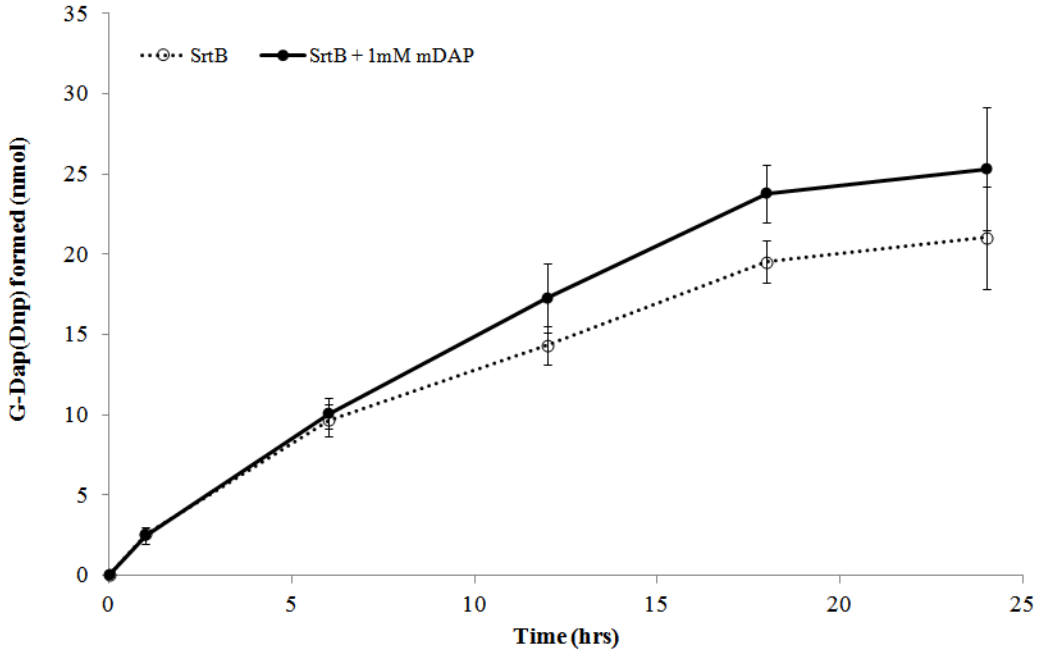


Figure 8.3.9 - Progress Curves of Cleavage of Abz-SPKTG-Dap(Dnp) by SrtB at 200µM. Mean values are plotted (n=3), with error bars indicating +/- 1SD.

At all SrtB concentrations, addition of 1mM mDAP resulted in a greater total quantity of G-Dap(Dnp) formed at 24hrs. At concentrations of 50µM and 100µM, formation of G-Dap(Dnp) was linear throughout the assay in the presence of mDAP, wheras in the absence of mDAP, activity reduced after around 8hrs, resulting in a flattening of the progress curve.

At the enzyme concentration of 200µM, a reduction in enzyme activity was observed after 6hrs in both the presence and absence of mDAP.

8.3.7 – Effect of Metal Ions and Salts on cleavage of Abz-SPKTG-Dap(Dnp)

As the *S. aureus* SrtB enzyme is calcium dependent, it was considered whether low activity of *C. difficile* SrtB was due to absence of a metal cofactor. To provide provisional information on the effect of various ions on activity of SrtB, Abz-SPKTG-Dap(Dnp) at a concentration of 250µM was incubated with 50µM SrtB in the presence of mDAP and various metal ions, EDTA or $(NH_4)_2SO_4$

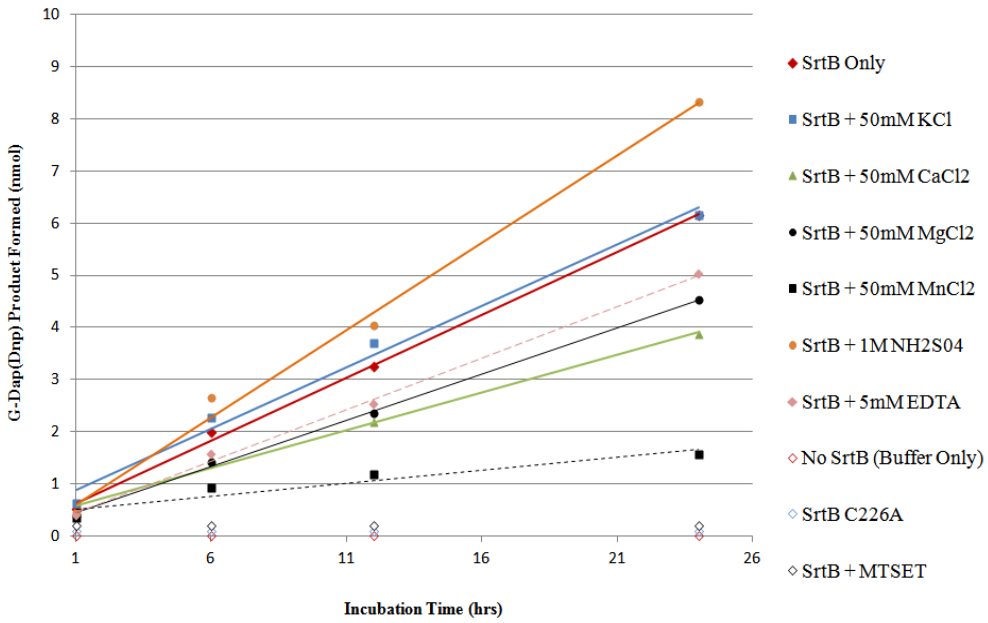


Figure 8.3.10 - Progress Curves of Abz-SPKTG-Dap(Dnp) Cleavage in the Presence of Various Salts. Trend lines are drawn for a linear fit.

Velocities of the reactions are shown in Table 8.3.3. Addition of any divalent metal salt reduced velocity, as did addition of EDTA, wheras addition of ammonium sulphate resulted in a modest increase. Addition of KCl did not affect velocity. Of the metal salts, $MnCl_2$ exhibited the greatest inhibitory effect.

Salt Added	Velocity (nmol.hr^{-1})	Velocity (% of SrtB only)
None	0.242	100%
50mM KCl	0.236	98%
50mM $CaCl_2$	0.145	60%
50mM $MgCl_2$	0.178	74%
50mM $MnCl_2$	0.050	21%
1M NH_2SO_4	0.334	138%
5mM EDTA	0.178	74%

Table 8.3.3 – Velocity of SrtB in the presence of various salts.

8.3.8 LC-MS analysis Mass Spectroscopy of SrtB Reactions

To provide mass spectroscopic confirmation of the products of the reaction between SrtB and the CWSS-mimetic peptides, reaction samples were analysed by LC-MS as described in Section 8.2.9. An index of samples analysed is shown in Table 8.3.4.

Rxn	Substrate	SrtB Conc (μM)	mDAP Conc (mM)	Figure
1	Abz-SPKTG-Dap(Dnp)	100	0	8.3.11
2	Abz-SPKTG-Dap(Dnp)	100	1	8.3.12
3	YSSSPKTGSSY	100	0	8.313
4	YSSSPKTGSSY	100	1	8.3.14

Table 8.3.4 – Summary of reactions analysed by LC-MS.

Annotated Extracted Ion Chromatographs (EICs) are shown for each of these reactions over the subsequent pages. Tabulated summaries of the EICs are shown in Tables 8.3.5 and 8.3.6.

Analyte	Mass	Detection of Analyte	
		Rxn 1	*Rxn 2*
Abz-SPKT	550	+	+
G-Dap(Dnp)	327	+	+
Abz-SPKTG-Dap(Dnp)	860	+	+
mDAP	190		+
Abz-SPKT-mDAP	722		+

Table 8.3.5 – Summary of EICs for reactions with Abz-SPKTG-Dap(Dnp).

Analyte		Detection of Analyte	
	Mass	*Rxn 3*	*Rxn 4*
YSSSPKTGSSY	1162	+	+
YSSSPKT	768	+	+
GSSY	412	+	+
mDAP	190		+
YSSSPKT-mDAP	940		+

Table 8.3.6 – Summary of EICs for reactions with YSSSPKTGSSY.

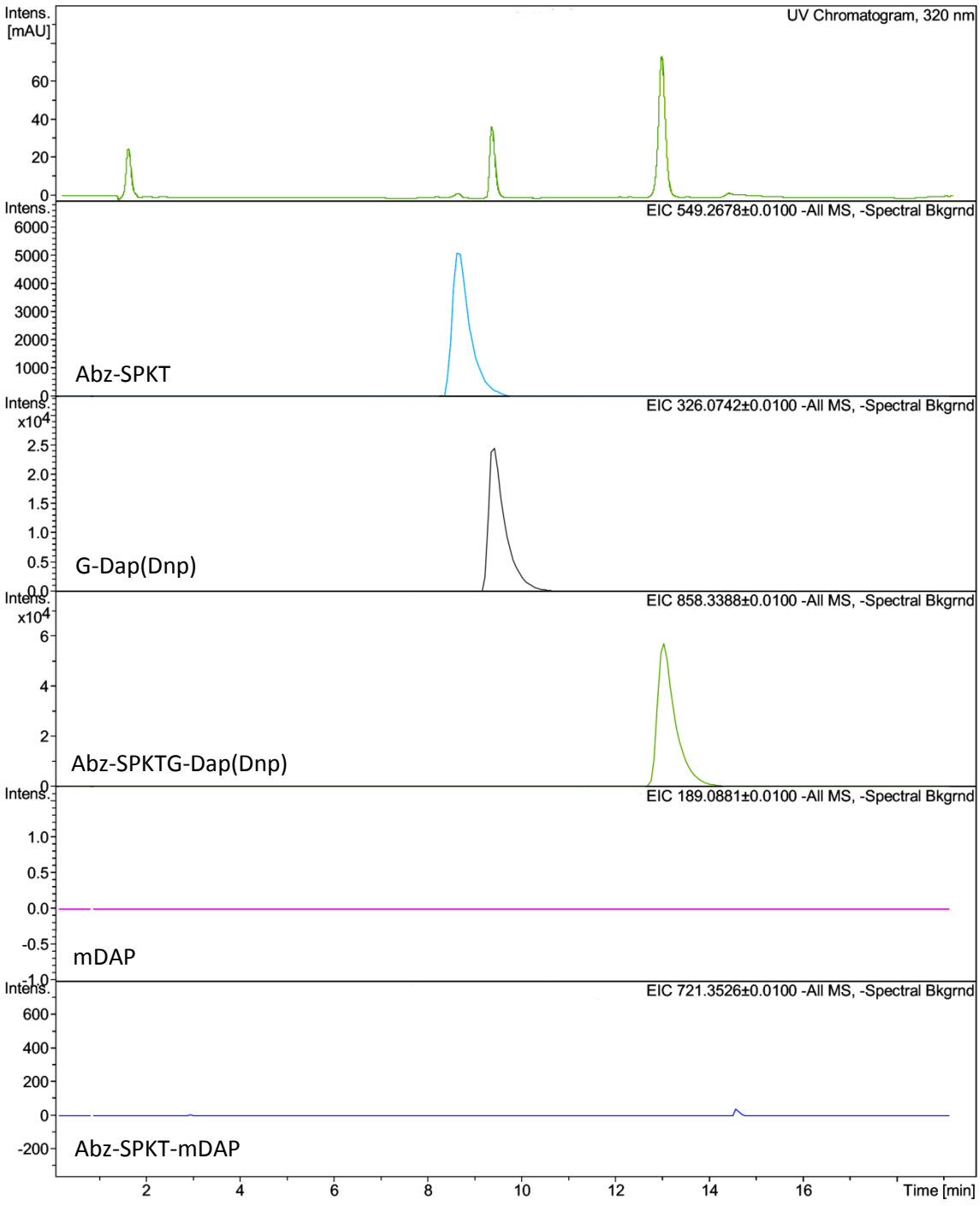


Figure 8.3.11 - Extracted Ion Chromatographs for Abz-SPKTG-Dap(Dnp) cleavage by SrtB in absence of mDAP. The UV chromatograph is shown (top) followed by EICs for each of the analytes. The mass of each analyte is shown at the top right of the EIC and its putative assignment at the bottom left. Note that intensity scales are not equivalent between EICs.

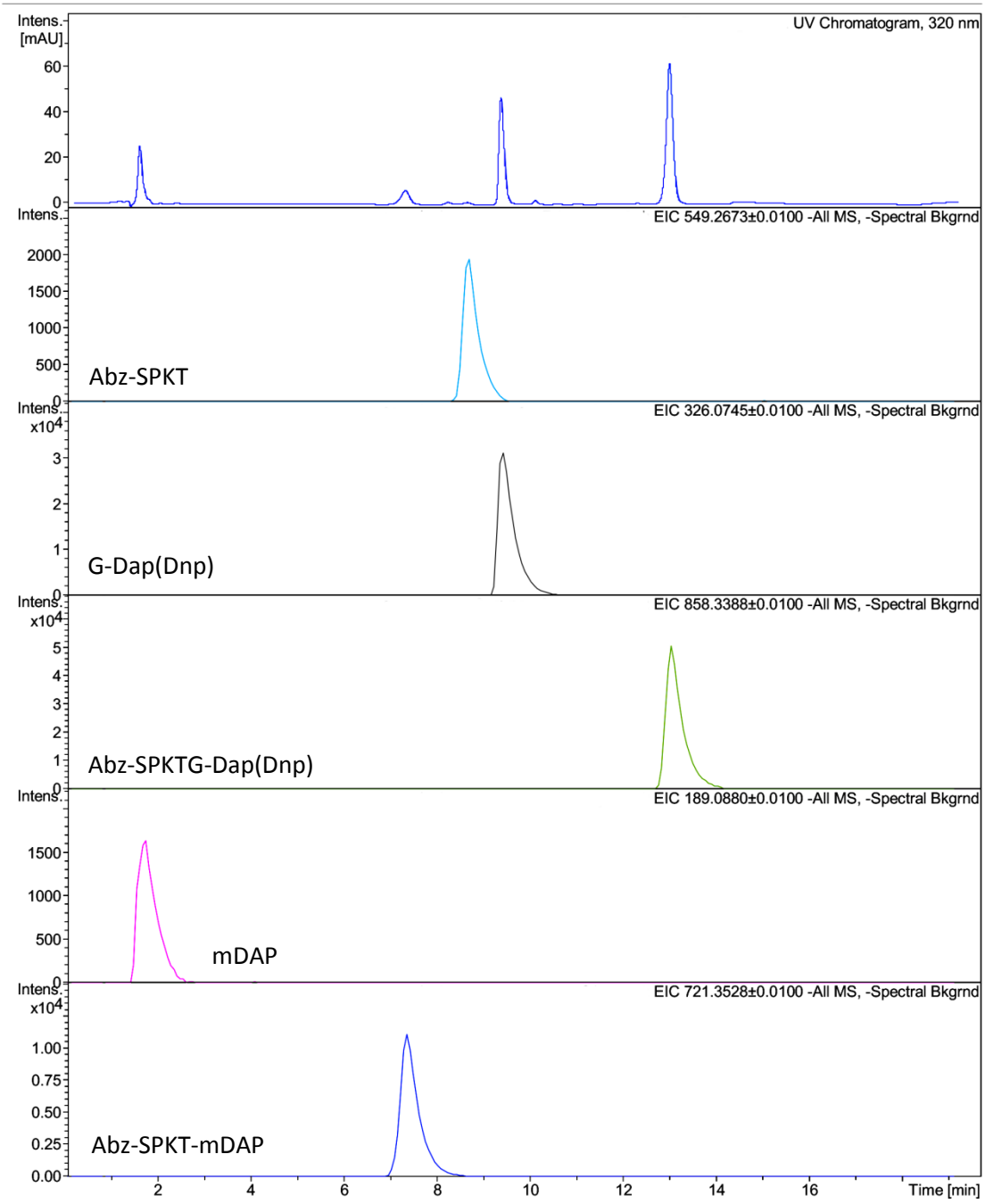


Figure 8.3.12 - Extracted Ion Chromatographs for Abz-SPKTG-Dap(Dnp) cleavage by SrtB in the presence of mDAP. The UV chromatograph is shown (top) followed by EICs for each of the analytes. The mass of each analyte is shown at the top right of the EIC and its putative assignment at the bottom left. Note that intensity scales are not equivalent between EICs.

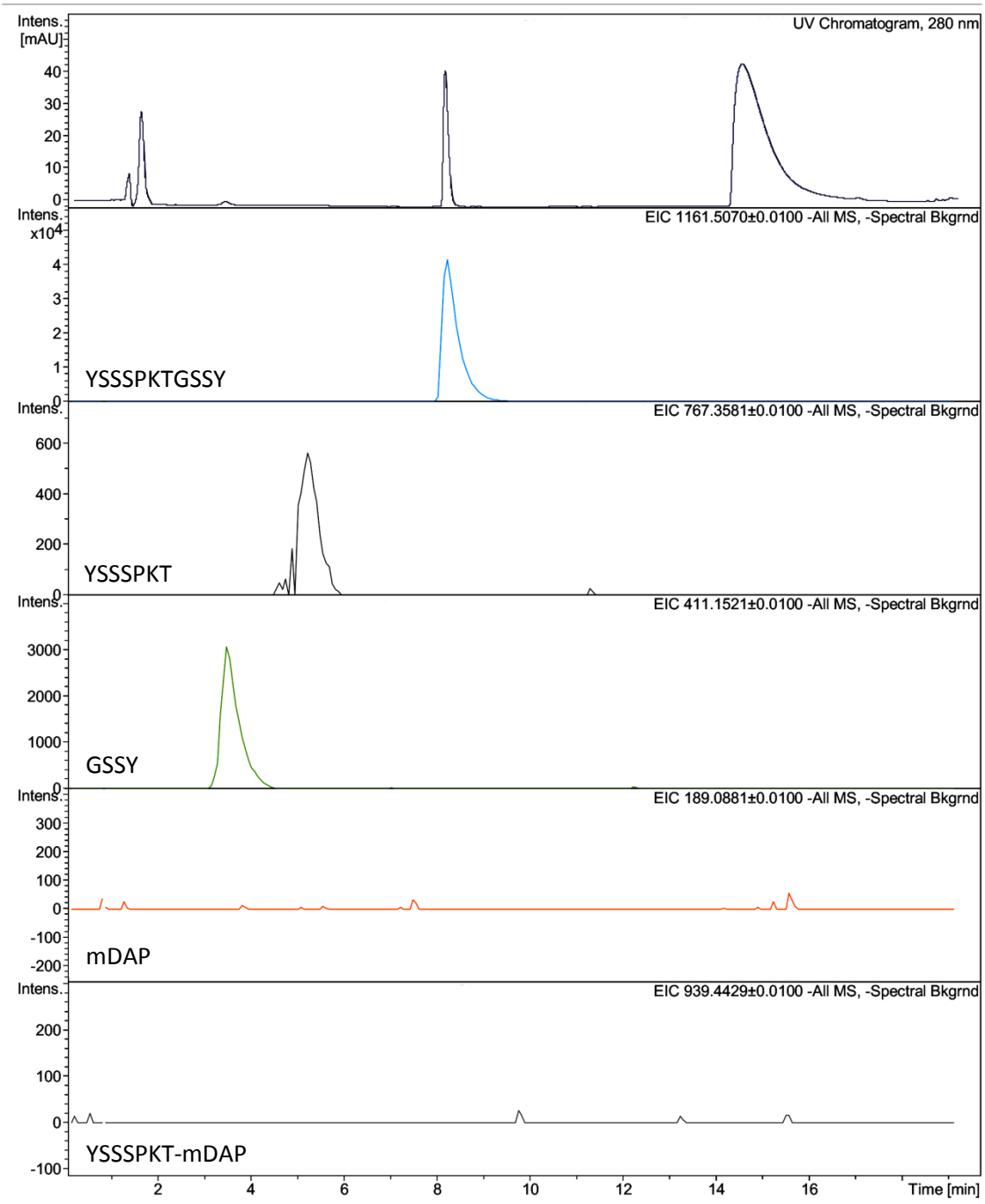


Figure 8.3.13 - Extracted Ion Chromatographs for YSSSPKTGSSY cleavage by SrtB in the absence of mDAP. The UV chromatograph is shown (top) followed by EICs for each of the analytes. The mass of each analyte is shown at the top right of the EIC and its putative assignment at the bottom left. Note that intensity scales are not equivalent between EICs.

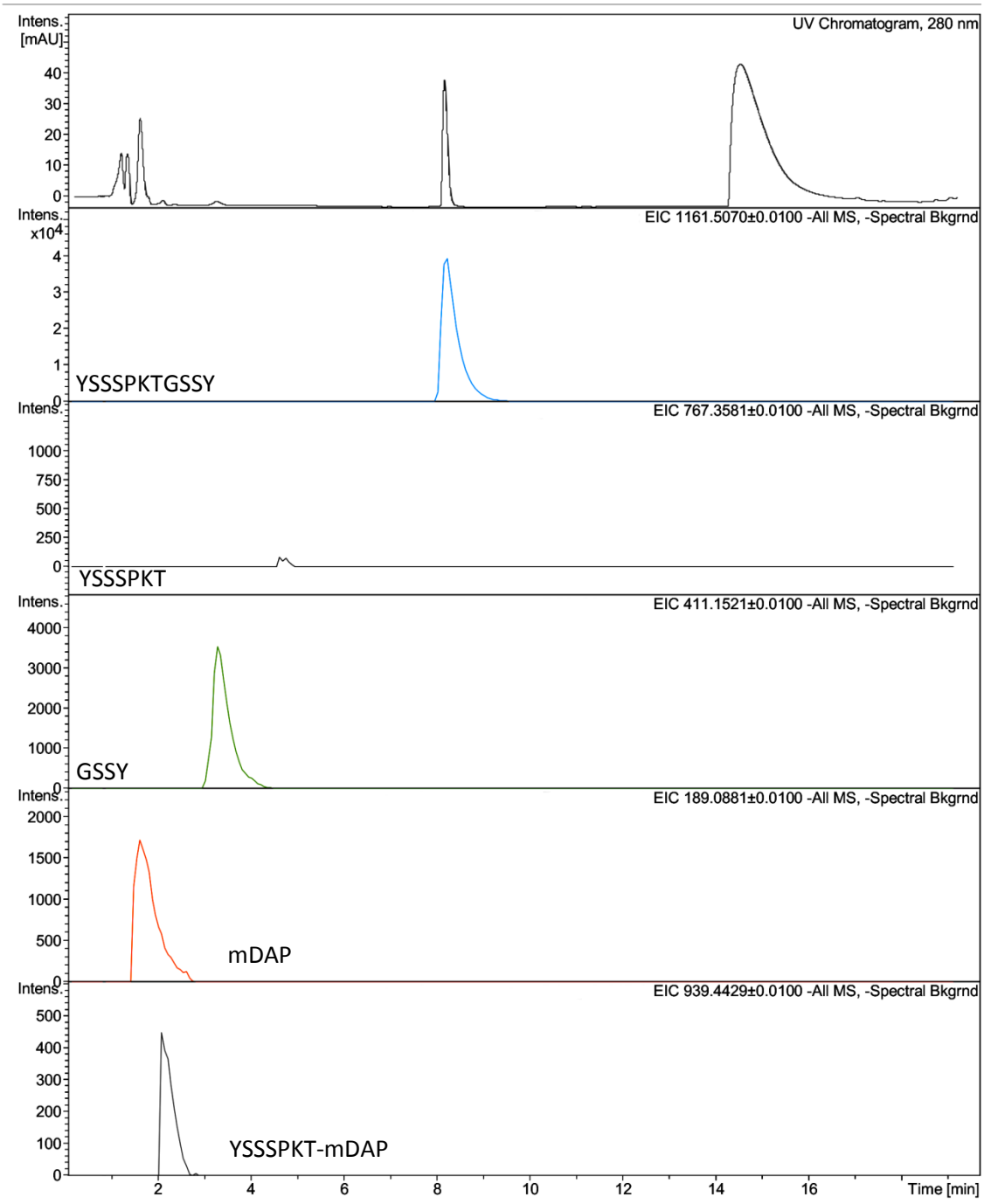


Figure 8.3.14 - Extracted Ion Chromatographs for YSSSPKTGSSY cleavage by SrtB in the presence of mDAP. The UV chromatograph is shown (top) followed by EICs for each of the analytes. The mass of each analyte is shown at the top right of the EIC and its putative assignment at the bottom left. Note that intensity scales are not equivalent between EICs.

8.3.9 - Identification of a SrtB Acyl Intermediate Species by UV-HPLC and LC-MS

Upon extended HPLC analysis of reactions between SrtB and Abz-SPKTG-Dap(Dnp), a 320nm absorbance peak was observed, eluting after Abz-SPKTG-Dap(Dnp) (Figure 8.3.15). This peak was significantly diminished when the reaction was performed in the presence of 1mM mDAP (Figure 8.3.16) and was not present in reactions where SrtB had been treated with the sulphydryl-reactive inhibitor MTSET (Figure 8.3.17). Analysis of these reactions while recording absorbance at 280nm in addition to 320nm and 370nm revealed that the 320nm absorbance co-eluted with the sortase enzyme as indicated by the sole absorbance peak at 280nm. As absorption at 320nm is characteristic of the Abz moiety of the probe, the absorption of the sortase enzyme at 320nm is consistent with the formation of a stable acyl intermediate (Abz-SPKT-SrtB).

To confirm this hypothesis, LC-MS spectra of reactions between SrtB and the two SPKTG peptides described in Section 8.2.3 were re-examined. For both peptides, the ion peak series corresponding to the mass of SrtB was accompanied by a minor peak series which was overlooked in the original automated analysis. Deconvolution of this minor peak series revealed a mass consistent with an acyl-enzyme (Figures 8.3.13).

Unfortunately, the poor absorbance of Abz precludes quantification of either the Abz-SPKT peptide or the acyl species, so it is not possible to estimate the proportion of enzyme in a stable acyl state. However, it is notable that for the SrtB reaction, the 320nm absorbance peak for the acyl intermediate is significant while that for Abz-SPKT is below the limit of detection, suggesting that ‘stalling’ of the reaction at the acyl state is favoured over the hydrolysis reaction. Addition of mDAP results in a reduction in size of the acyl peak, with it being roughly equal in area to that of Abz-SPKT.

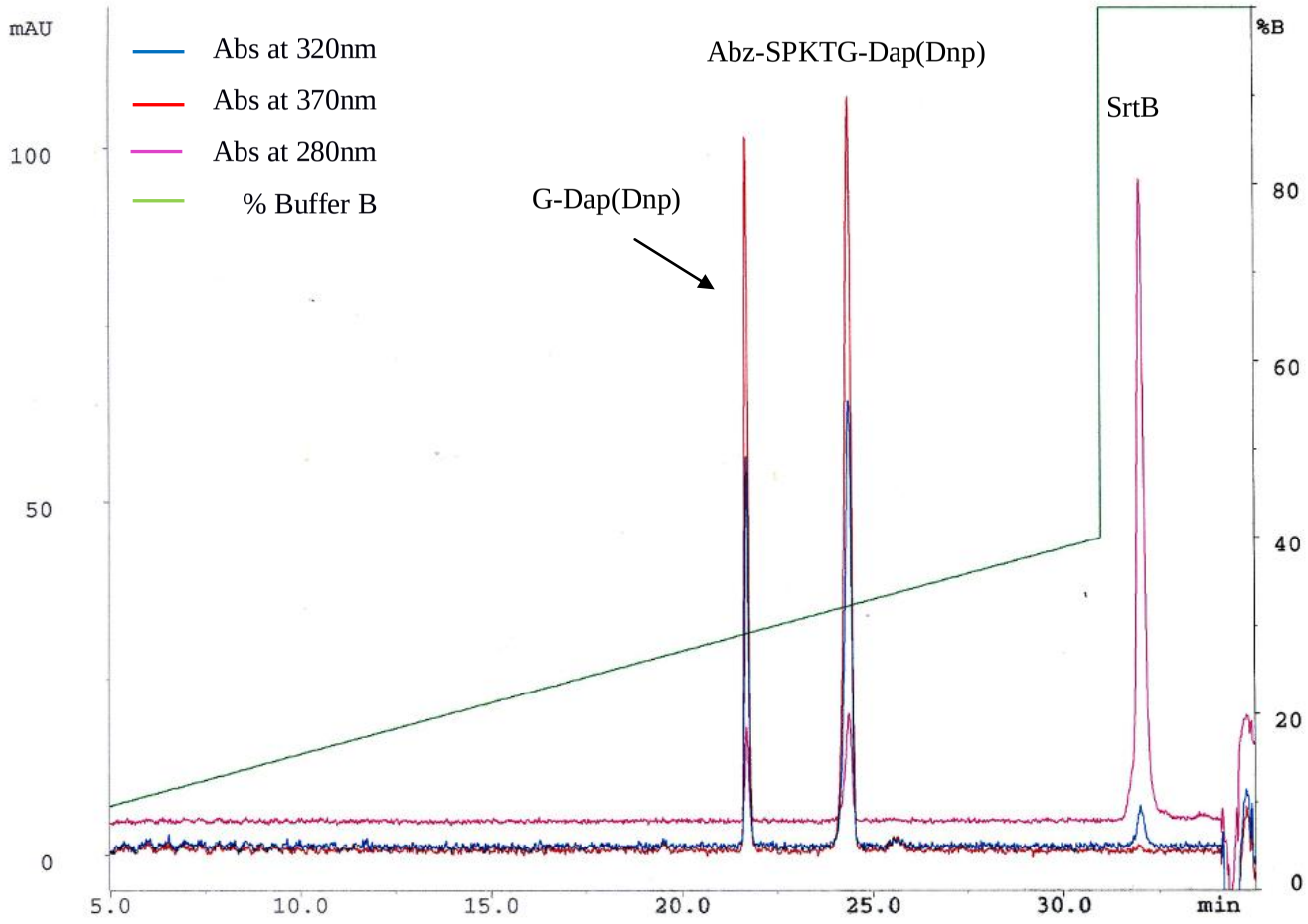


Figure 8.3.15 – HPLC analysis of reaction between SrtB and Abz-SPKTG-Dap(Dnp). The large peak of absorbance at 280nm corresponds to elution of SrtB. Co-elution of this analyte with a peak of 320nm absorbance is clearly visible.

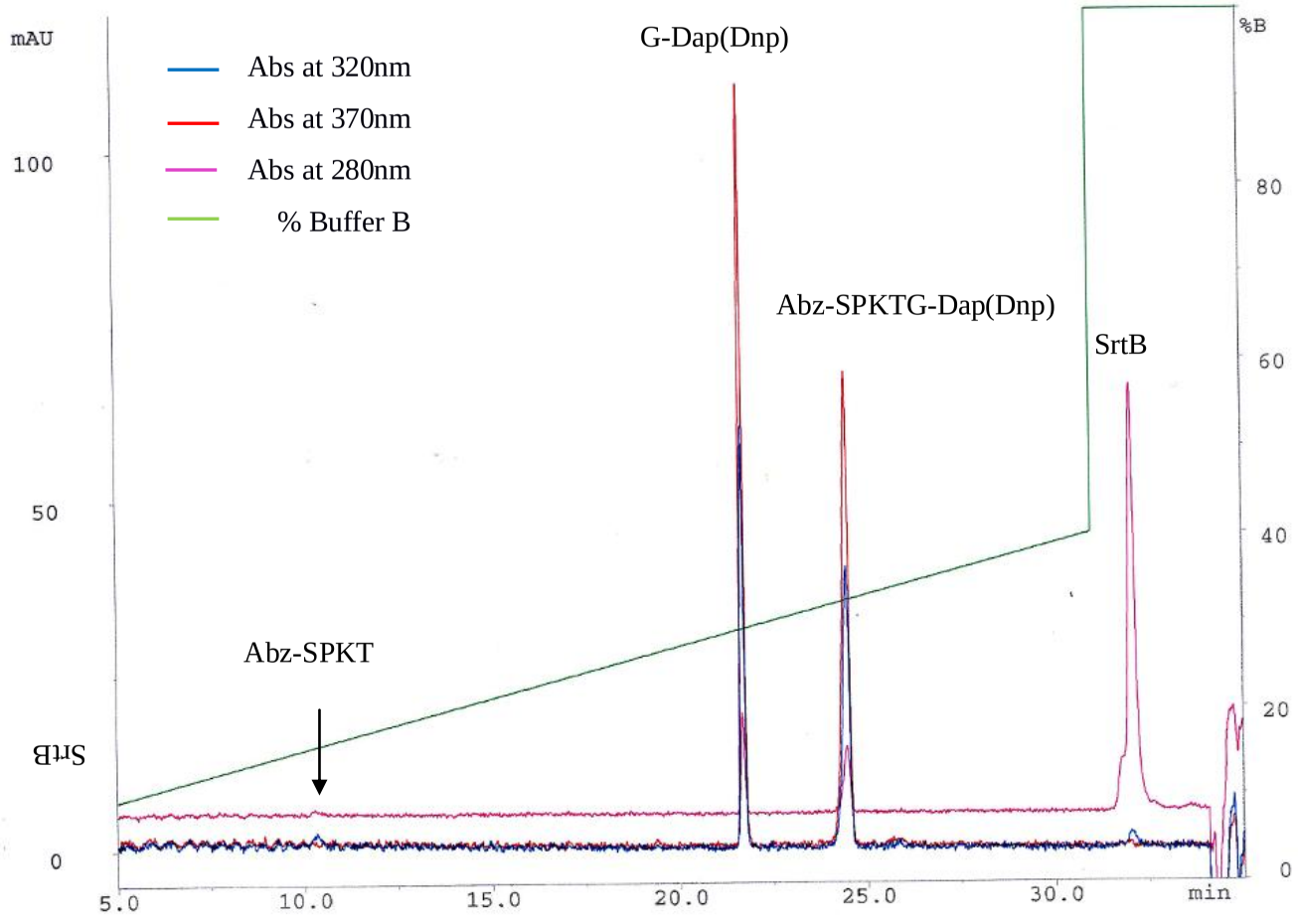


Figure 8.3.16 – HPLC analysis of reaction between SrtB and Abz-SPKTG-Dap(Dnp) in the presence of 1mM mDAP. The 320nm absorbance peak associated with elution of SrtB is present, but diminished relative to that observed in the absence of mDAP.

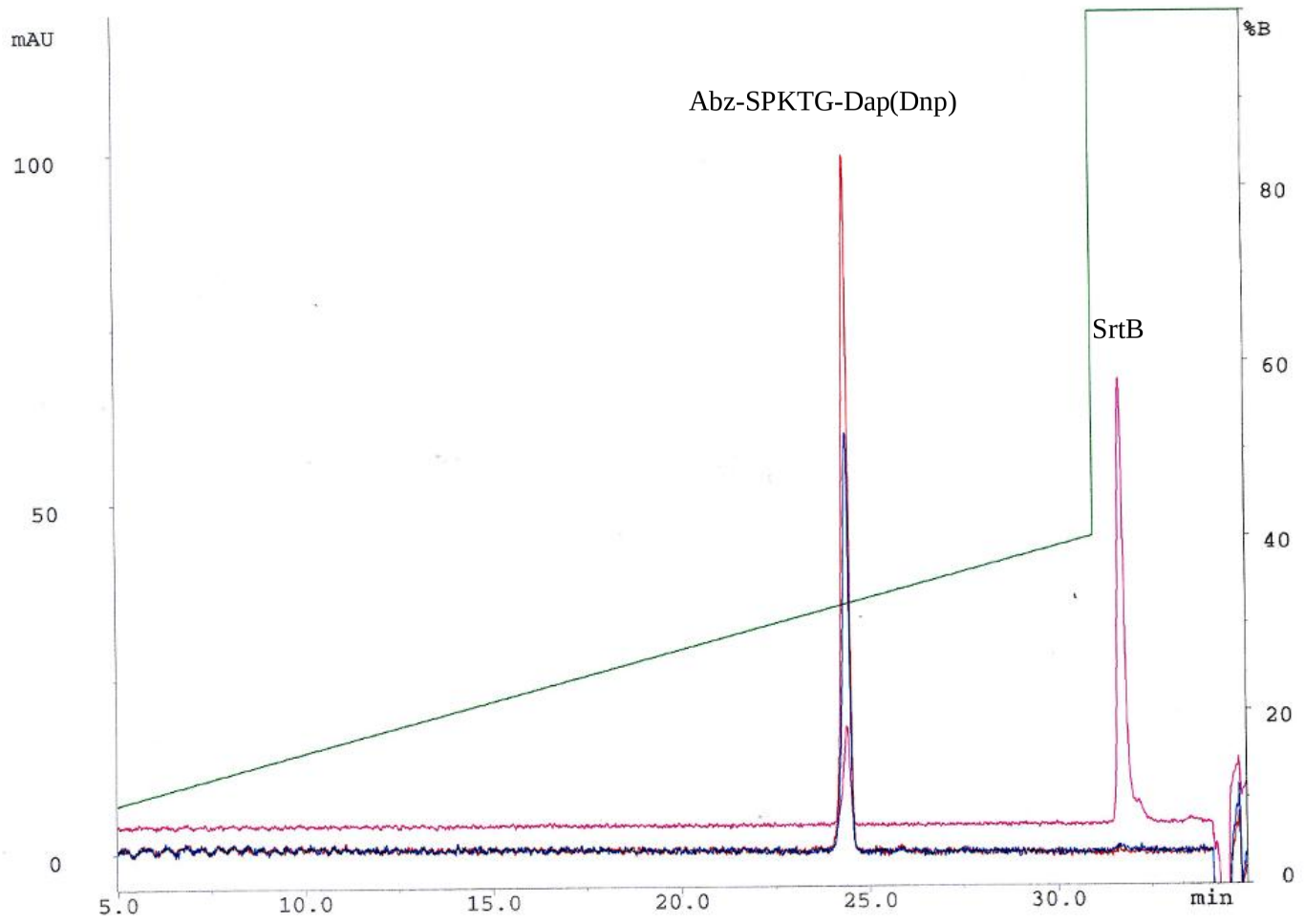
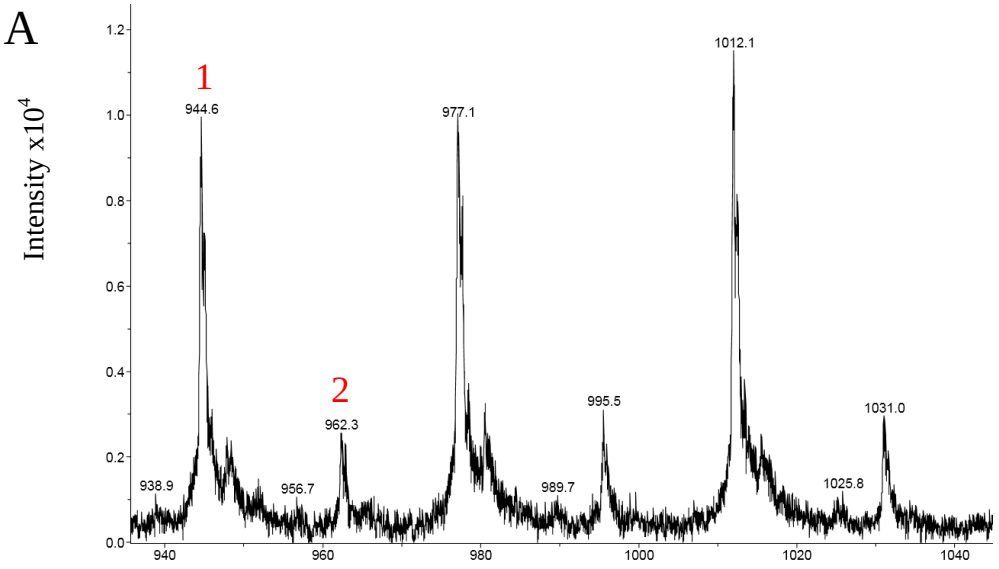
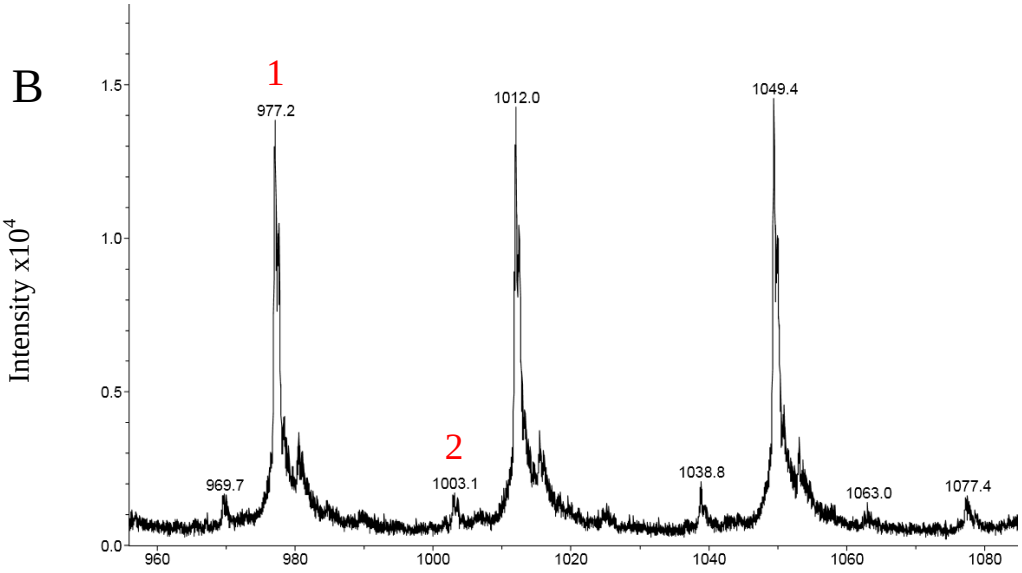


Figure 8.3.17 – HPLC analysis of reaction between SrtB and Abz-SPKTG-Dap(Dnp) in the presence of the sulphydryl reactive inhibitor MTSET. Elution of SrtB is not associated with a co-eluting peak of 320nm absorbance.



Peak	Assignment	Theoretical Mass	Observed Mass
1	SrtB	28308	28306
2	Abz-SPKT-SrtB	28840	28838



Peak	Assignment	Theoretical Mass	Observed Mass
1	SrtB	28308	28307
2	Abz-SPKT-SrtB	29058	29055

Figure 8.3.18 - LC-MS detection of Acyl SrtB. Positive mode mass spectra of SrtB after reaction with (A) Abz-SPKTG-Dap(Dnp) and (B) YSSSPKTGSSY. Results of peak deconvolution are shown in the accompanying tables.

8.4 - Discussion

8.4.1 Preliminary Characterisation of SrtB activity by a FRET-based Assay

The cleavage of SPKTG, TKSGP and NVQTG peptides was assessed using a FRET based assay. The C226A mutant was employed as a negative control, as this mutation has been demonstrated to abolish all activity in other sortases (Mazmanian *et al.*, 1999). A statistically significant increase in fluorescence over the course of the assay was observed only when the SPKTG peptide was incubated with SrtB. For other peptides and for an SPKTG peptide incubated with SrtB C226A, modest increases were observed, presumably due to degradation of the probe during incubation period.

The error in the FRET assay was very high, with standard deviations approaching the value of the mean for samples in which no specific cleavage was detected. This is likely attributable to the high gain values required in quantification of fluorescence, resulting in a low signal:noise ratio. With the benefit of knowledge from later quantitative experiments, it is clear that the requirement for high gain values was due to the poor activity of the enzyme resulting in minimal product formation. As the assay was intended to be purely qualitative, no efforts were made to quantify the extent of cleavage. This could be achieved by preparation of a standard curve, but due to inner filter quenching effects discussed in Section 8.1.1, such a standard curve would not be reliable for quantification.

Despite the severe shortcomings described above, the data are suggestive that SrtB cleaves SPKTG but not NVQTG or TKSGP, and that SrtB C226A is inactive as expected. To confirm these results in a more robust manner and to quantify product without interference from inner filter effects, the FRET assay was abandoned in preference for a HPLC-based assay.

8.4.2 Assay of SrtB Activity by the HPLC method

In order to provide quantitative data on the cleavage of peptides by SrtB, a HPLC-based SrtB assay was established, with UV detection enabled by modification of the SPKTG and NVQTG peptides with Abz and Dap(Dnp) at the N- and C- termini respectively. To allow robust quantification of the reaction participants, a set of standards was synthesised corresponding to the products of cleavage of the probe at the threonyl-glycl bond.

The ‘long’ HPLC method described in Section 8.3.2 resolves both of the hydrolysis products plus residual intact substrate. However, the poor absorption characteristics of Abz-SPKTG render its quantification impractical, so resolution of this peak is not strictly necessary. A shorter method was therefore developed which would allow more rapid quantification of G-Dap(Dnp) and Abz-

SPKTG-Dap(Dnp) only. Progress of the reaction could therefore be monitored either by appearance of the G-Dap(Dnp) or disappearance of Abz-SPKTG-Dap(Dnp). The elution order of standards was Abz-SPKT, G-Dap(Dnp) then Abz-SPKTG-Dap(Dnp), indicating that the Dap(Dnp) moiety was the major contributor to retention rather than the peptide probe. As shown in Figure 8.3.5, detection of Abz-SPKTG-Dap(Dnp) was linear over the range from 100-1000µM and accuracy of injection was also good. Cleavage of an SPKTG but not NVQTG motif confirms the predictions of Comfort and Clubb (2004) and disproves the speculative suggestion of Tulli *et al* (2013) that the *C. difficile* collagen adhesin CbpA may be anchored by the *C. difficile* sortase via its NVQTG motif.

Upon observation of poor hydrolysis of Abz-SPKTG-Dap(Dnp), it was proposed that the poor activity was due to steric hindrance by the relatively bulky modifications at the N- and C-termini. A new set of peptides was therefore synthesised, with diserine spacers at either side of the tetrapeptide CWSS, and tyrosine residues at the N- and C- termini to allow detection by UV spectroscopy at 280nm. However, as shown in Figure 8.3.7, hydrolysis of this peptide was even poorer than of Abz-SPKTG-Dap(Dnp). The Ser-Ser-Tyr is a significantly polar moiety relative to the aromatic Abz and Dnp, this may indicate that Abz-SPKTG-Dap(Dnp) participates in hydrophobic interactions with SrtB outside of the core pentapeptide motif.

At sortase concentrations of 50µM and 100µM, reactions progress in a linear fashion throughout the assay in the presence of mDAP, whereas in the absence of mDAP product formation departs from the linear trend after around 6 hrs. Initial rates are similar regardless of whether mDAP is present or absent, but by 24hrs, the extent of cleavage is significantly greater in the presence of mDAP. Flattening of enzyme progress curves (reflecting a reduction in reaction rate over time), is generally ascribed to either exhaustion of substrate or inactivation of the enzyme. Since it is known that the substrate was in excess throughout the experiment, it is more likely that in this case it is due to changes in the enzyme. Taken in combination with the evidence of acyl intermediate formation presented in Section 8.3.8, it is tentatively suggested that the observed differences in activity in the later phases of the assay are due to accumulation of acyl (and therefore inactive) SrtB in the absence of mDAP.

Given the observation of *in vitro* poor activity of SrtB (which will be discussed in Section 8.4.5), SrtB activity was assayed in the presence of metal ions to exclude the possibility of dependence upon metal ions for activity, as is the case in *S. aureus* SrtA. The opposite effect was observed, with SrtB inhibited by divalent cations. This result is consistent with those of Race *et al* (2009), who observed that the *S. pyogenes* SrtA was inhibited by Ca^{2+}. In that case, the authors did not report testing of other metal ions or investigate the phenomenon any further. Given that the active

site of sortase enzymes contains a conserved histidine, a tentative mechanism for inhibition could be aberrant ion coordination distorting the active site. To determine whether SrtB does bind metal ions, a technique such as isothermal titration calorimetry could be employed.

8.4.3 Discussion on the Poor Activity of SrtB

Poor activity of sortase enzymes has been widely commented upon in the literature. Indeed, assay of *S. pyogenes* SrtB by a FRET based method failed to detect any *in vitro* activity at all (Kang *et al.*, 2011). Wiener *et al* (2010) detected cleavage of a CWSS peptide by *B. anthracis* SrtB, but a transpeptidation reaction with mDAP could not be detected. These observations of poor activity are significantly at odds with observations of sortase activity *in vivo*. Pulse-chase experiments indicate that in *S. aureus*, SpA is anchored to the cell wall by SrtA within 2-3 min of its expression (Mazmanian *et al.*, 1999). More generally, it is unlikely that an enzyme with such a low turnover rate would be of any benefit to an organism with a generation time of around 30mins.

To understand this disparity between *in vivo* and *in vitro* activity, it is necessary to consider that the *in vitro* assays are very poor simulations of what occurs *in vivo*. Most importantly, *in vitro* sortase assays are performed in solution, wheras *in vivo* sorting is surface localised by definition and appears to occur at distinct foci co-localised with the Sec apparatus (Hu *et al.*, 2008; Raz and Fischetti, 2008). This co-localisation of enzyme and substrate increases effective substrate concentrations in a way which is poorly represented by experiments performed in solution.

Another consideration is the suitability of the substrates used for *in vitro* assays. A short peptide mimicking the CWSS may not be a realistic surrogate for the CWSS motif of wall-sorted substrate, which is part of a large protein, with the associated secondary and tertiary structure. Interactions of sortase enzymes with their substrates may be significantly more extensive than is currently known, with current structural knowledge of enzyme-substrate interactions limited to that gleaned from structures of sortases complexed with the same short peptides as used for kinetic studies. Indeed, the differences in SrtB activity towards YSSSPKTGSSY and Abz-SPKTG-Dap(Dnp) reported in this work indicate that interactions outside of the pentapeptide CWSS motif may be important for substrate recognition.

It is notable that for convenience of soluble expression, the characterisation described in this work (in common with all published characterisations of sortase activity), utilised a truncated enzyme omitting its N-terminal hydrophobic anchor region. Wiener *et al* (2010) investigated kinetic properties of two variants with truncations of varying length, discovering that N-terminal truncation of the enzyme to remove the unstructured region before the short H1 increased k_{cat} by 20%. To explain this phenomenon, the authors offered a hypothesis that the N-terminus may

transiently occupy the active site and prevent substrate binding. However, this was in a class A sortase where a the absence of a long H1 helix places the N-terminus much closer to the active site than in a class B sortase. Given its spatial separation from the active site, it seems unlikely that the N-terminus of a class B sortase would significantly affect enzyme activity independently of its role in surface localisation.

It is quite possible that the particularly poor performance of *C. difficile* SrtB described in this work is due specifically to the method by which it was prepared. Loss of activity during purification is a well documented phenomenon, and can occur for many reasons including aggregation, exposure to high temperatures and exposure to inhibitory compounds.

The apparent inhibition of SrtB by divalent cations is a particular concern given that the purification method included a nickel affinity step. While addition of EDTA did not increase enzyme activity (indicating that no inhibitory metal ions were bound to the enzyme) it is still of concern that purification of the enzyme exposes it to high concentrations of a potentially inhibitory molecule. It would be therefore be worthwhile to assay SrtB enzyme purified by a different method. Hydrophobic interaction chromatography would be worthy of investigation as a first chromatographic step, as it performs very well as a second step in the purification of SrtB as described in Chapter 4.

8.4.4 LC-MS analysis of SrtB products

Due to its very strong ion suppression activity, TFA is not suitable as a mobile phase buffer for chromatography where mass spectroscopy is to be performed. However, MS compatible buffers such as acetic acid and formic acid often result in poor chromatography due to low ion pairing performance. In this experiment, replacement of TFA with formic acid did not impair resolution.

The first aim of the LC-MS analysis was to confirm the identity of the cleavage products identified in the HPLC assay and hence confirm that cleavage occurred at the thronyl-glycyl bond of the peptide, as would be expected from a sortase-mediated hydrolysis reaction. Comparison of elution profiles with those of the standards was not sufficiently robust because, as previously discussed, the major contribution to retention appears to be the modifications rather than the peptides, so cleavage at a different position in the peptide may not significantly alter the elution profile of the probe. The observed product masses corresponded unequivocally with the theoretical masses of the expected cleavage products, confirming that cleavage was at the threonyl-glycyl bond.

The second aim of the LC-MS analysis was to detect transpeptidation products resulting from reaction of the acyl enzyme with meso-diaminopimelic acid (mDAP). This was a key experiment in

order to definitively characterise *C. difficile* SrtB as a true sortase capable of a full transpeptidation reaction. When either of the SPKTG peptides was incubated with SrtB in the presence of mDAP, mass spectroscopy indicated the presence of species with masses corresponding to the expected transpeptidation products YSSSPKT-mDAP (939 Da) and Abz-SPKT-mDAP (721 Da). Notably, the hydrolysis products 'Abz-SPKT' and YSSSPKT were still present in reactions where mDAP was in molar excess, indicating that hydrolysis occurs even where a transpeptidation substrate is available. This may reflect the unavoidable fact that due to the reaction being performed in an aqueous solution, water was always in excess relative to mDAP. *In vivo*, hydrolysis is an undesirable outcome for the bacterium because it results in secretion, and hence loss, of proteins which were intended for surface anchoring. However, in the tightly surface-localised environment in which sortases operate, it is likely that the effective concentration of the transpeptidation substrate is sufficiently high to favour transpeptidation over hydrolysis.

8.4.5 Identification of an Acyl SrtB Intermediate

Although there are several proposed models for the reaction mechanism of sortase enzymes (as reviewed in Chapter 2), all of these mechanisms lead to a covalent intermediate state. Cleavage of the CWSS peptide at the threonyl-glycyl bond creates an acyl enzyme, with a thioester bond formed between the active site cysteine of the enzyme and the carboxyl of the P4 threonine. Previously, it has been demonstrated that this acyl state is detectable by mass spectroscopy (Huang *et al.*, 2003).

In this work, both chromatographic and mass spectroscopic evidence is presented for an acyl complex between *C. difficile* SrtB and an SPKT peptide. It is also observed that in the absence of the transpeptidation substrate mDAP (where hydrolysis reaction is the only means of returning the enzyme to its resting state), the extent of acylation is qualitatively higher. This may indicate that mDAP is a better nucleophile than water with regard to attack on the acyl enzyme. However, the generally poor performance of SrtB prevents any firm conclusions from being drawn, as it is clear that the enzyme is not active in a physiologically relevant manner.

8.4.6 Limitations and Further Work

The work described in this chapter is subject to several limitations, most of which result from the low *in vitro* activity of the enzyme. Most notably, it was not possible to determine standard enzymatic parameters such as Vmax and K_M; doing so would have consumed vast quantities of enzyme and would have been entirely inappropriate, since it is clear that the *in vitro* activity of the enzyme is seriously impaired and unlikely to bear any relevance to its *in vivo* activity. On a more practical level, difficulties and delays encountered in preparing sufficient active SrtB resulted in the provisional kinetic data being collected in a short space of time. Without access to an automatic

sample loader and with an analysis time of around 30mins including sample preparation, the amount of work which could be performed in the available time was very limited. The use of higher performance chromatography methods could reduce analysis time significantly – Kruger *et al* (2004) report an analysis time of 5mins.

Poor absorption by the aminobenzoic acid moiety which tagged the N-terminus of the CWSS peptide prevented quantification of either the Abz-SPKT hydrolysis product, the Abz-SPKT-mDAP transpeptidation product or the Abz-SPKT-SrtB acyl enzyme. This could be effected by replacement of Abz with a more strongly absorbing chromophore – alternatively, a fluorescence detection method for aminobenzoic acid could be employed, as has been described previously (Kruger *et al.*, 2004c).

Future work must focus on determining whether the low activity exhibited by SrtB in this work is an inherent property of the enzyme when *in vitro*, due to reasons described in Section 8.4.5, or whether the enzyme is becoming inactivated during purification. Proper kinetic characterisation may then be performed for comparison with other sortase enzymes.

8.5 - Summary

Due to the low activity of the enzyme as well as practical limitations, kinetic information described in this chapter is limited in its scope. However, development and evaluation of the highly sensitive HPLC and LC-MS methods as described in this chapter will be of great value for further investigation of the reaction.

The results presented in this chapter support the hypothesis that the *C. difficile* SrtB enzyme cleaves an SPKTG motif rather than an NVQTG motif, and that it is capable of a transpeptidation reaction with meso-diaminopimelic acid (mDAP), the diamino acid which forms the cross links of the *C. difficile* peptidoglycan. Additionally, an acyl enzyme intermediate state was identified by chromatographic analysis and confirmed by mass spectroscopy.

Chapter 9 – General Discussion

9.1 *Clostridium difficile* – Current Situation and Future Challenges

Clostridium difficile is an archetypal opportunistic pathogen, generally living within us as a harmless commensal organism and only causing disease in very specific circumstances. The iatrogenic depletion of gut flora which leads to *C. difficile* infection has only occurred on a large scale very recently, as an ironic consequence of the development of potent broad-spectrum antibiotics. While *C. difficile* was well known to clinicians from the 1970s onwards, it was not until the dramatic change in its epidemiology in the early 2000s that it became a pathogen of widespread public concern. The notoriety of nosocomial pathogens and their effect on the confidence of patients cannot be underestimated - a poll of 1,040 people in 2008 revealed that hospital infections were the primary concern of NHS patients, with 40% of respondents most concerned about infection risk, more than those concerned about waiting lists, understaffing or poor quality of care (ICM/BBC, 2008).

While it is encouraging to observe the current downward trend in *C. difficile* cases since the peak epidemic years (Figure 1.1.3), there is good reason to remain vigilant. There are growing reports of *C. difficile* infection in the wider community, including those with no exposure to antimicrobial therapy and in persons of much lower age than traditionally associated with nosocomial *C. difficile* infection (Khanna *et al.*, 2012; Kutty *et al.*, 2010). *C. difficile* cannot, therefore, be regarded purely as a healthcare associated infection. Just as its epidemiology changed abruptly with the introduction of clindamycin in the 1970s, future changes in human health may precipitate similar shifts in susceptibility. Even without expanding its reach into new populations, *C. difficile* is likely to be a growing problem over future decades due to the continuing expansion of the traditional at-risk populations, most notably the elderly and particularly those requiring prolonged hospitalisation or residential care.

One of the cornerstones of the management of *C. difficile* outbreaks is 'antibiotic stewardship', the reservation of certain antibiotics for specific infections and heavy regulation of the fluoroquinolones and third generation cephalosporins. However, *C. difficile* also poses a major challenge to the concept of antibiotic stewardship, as vancomycin is an antibiotic of last resort for *Enterococcus sp*. and *Staphylococcus sp*. yet it is the 'gold standard' for treatment for *C. difficile* infection. This is a particular concern given that patients at risk of *C. difficile* are also more likely to be at risk of developing other nosocomial infections due to advanced age and higher rates of co-morbidity (Loo *et al.*, 2011).

Current therapies for *C. difficile* are inadequate. While effective in killing *C. difficile*, vancomycin and metronidazole both suppress the glut flora even further, a 'fighting fire with fire' approach that leads to multiple relapses, ongoing misery for patients and serious financial consequences for healthcare providers; it has been estimated that 94% of the cost of *C. difficile* infection is related to increased duration of hospital stay (Wilcox *et al.*, 1996). The recently introduced antibiotic Fidaxomycin is a promising development, especially with regard to prevention of relapse, but it continues to rely upon the principle of treating an antibiotic-associated infection with further antibiotics. The case for novel, non-antibiotic therapies is perhaps stronger for *C. difficile* than for any other organism.

9.2 The Imperative for Novel Therapeutics

Arguably one of the greatest achievements of the 20th century was the conquering of bacterial pathogens through improved sanitation, vaccination and development of antibiotics. To the casual observer in the late 20th century, it could have appeared that bacteria were no longer a threat to human health in the developed world. With the exception of rapidly fulminating infections such as those of *B. anthracis* and *N. meningitidis*, there are few infections which cannot be successfully treated with currently available antibiotics. As well as their therapeutic applications, the prophylactic use of antibiotics has facilitated countless other advances in medicine, allowing development of modern surgery and the treatment of patients who are immunocomprimised.

Despite this apparent progress, we are undoubtedly heading towards a major microbial challenge in the 21st century. Of the 13 classes of antibiotics, all but four were discovered before 1960. While the discovery of antibiotics has stalled, the evolution of antibiotic resistance in bacteria has continued unabated, rapidly depleting the treatment options available to clinicians. It is quite remarkable that in a period of less than 100 years, a minute fraction of human history, humankind has conquered bacterial pathogens, only to face a return to the widespread occurrence of untreatable infections, the so-called 'post-antibiotic era'. The first portents of this era are already upon us, with the emergence of several pathogens which are resistant to their traditional 'last resort' antibiotic, among them the Vancomycin Resistant *Enterococcus spp.* (VRE), the Carbepenam Resistant *Enterobacteriaceae* (CRE) and Totally Drug Resistant *M. tuberculosis (TDR-TB)* (Courvalin, 2006; Migliori, 2007).

With regard to *C. difficile*, acquisition of antibiotic resistance presents unusually complex consequences. Not only does resistance to a particular antibiotic rule out its use for treatment of CDI, it can also trigger CDI in *C. difficile*-carrying patients who are given the antibiotic for treatment of an unrelated infection. Indeed, increasing resistance to the quinolones is one of the

factors implicated in the changes in *C. difficile* epidemiology in the early 2000s, including the emergence of the epidemic 027 strains, which are quinolone resistant (Pepin *et al.*, 2005a).

While the phenomenon of declining antibiotic discovery is often cited as an example of market failure, with pharmaceutical companies preferring development of more profitable drugs, a downward trend in discovery has been observed across all classes of drug in spite of year-on-year increase in research spending (Powers, 2004). It is therefore simplistic to assume that the solution to the antibiotic crisis is to increase funding for traditional antibiotic discovery programmes. New approaches to the problem are clearly required in *addition* to intensification of current efforts.

The majority of existing antibiotics target either cell wall synthesis, protein synthesis or nucleic acid synthesis, and as these antibiotics were mostly developed from secondary metabolites of other microorganisms, the choices of target were largely a result of nature. In the era of genomics, proteomics, systems biology and high throughput structural biology, the potential exists to reverse engineer the metabolic and pathogenic processes of bacteria and create 'designer antibiotics' based upon novel targets that nature may have overlooked. Strutural biology provides particular opportunities for development of 'designer drugs' with the emerging technology of structure-based drug design.

The principle of targeting bacterial virulence rather than bacterial metabolism is currently limited to treatment of toxin-mediated disease – anti-toxin therapy is routinely used for treatment of botulinum toxin poisoning, was used historically for the treatment of *Clostridium tetani* infection and has been demonstrated for the experimental treatment of *C. difficile* infection (Roberts *et al.*, 2012). However, there have also been several experimental demonstrations of therapies targeting non-toxin virulence factors. For example, the small molecule Virstatin has been shown to protect infant mice from infection with *Vibrio cholera* by inhibiting the dimerisation of a transcriptional regulator ToxT and consequently inhibiting the expression of the toxin-coregulated pili (Hung *et al.*, 2005; Shakhnovich *et al.*, 2007). In another example, inhibitory autoinducing peptides which interfere with quorum sensing in *S. aureus* have been shown to inhibit abscess formation in a mouse model (Wright *et al.*, 2005). It is noteworthy that even if these strategies prove to be poorly efficacious in comparison with current antibiotic therapies, virulence factors may seem much more attractive as therapeutic targets in the post-antibiotic era. For example, virulence-targeted therapies could be used for mild infections, while the antibiotics which remain effective could be reserved for severe infections.

One of the most interesting yet challenging aspects of targeting bacterial virulence is that such therapies would not be bactericidal, but merely inhibit the organism's ability to cause disease. A

considerable advantage of this is that a drug which does not kill bacteria is unlikely to exert a strong selective pressure for resistance (Since the pathogen gains considerable advantage from inhabiting the host, it would be incorrect to assume that there would be *no* selective pressure). Unfortunately, merely inhibiting pathogenesis leaves the patient's immune system to clear the infection. This is a particular problem because many infected patients are immunocomprimised, either as a consequence of their infection or due to their underlying poor health. Again, it is likely that bactericidal antibiotics would be preferred for such patients.

9.3 Sortase Enzymes as Targets of Novel Therapeutics

Sortase enzymes offer an intriguing target for treatment of bacterial infection by inhibition of virulence. One of the difficulties of targeting virulence is that it is often highly multifactorial – *S. aureus* is an extraordinary example of this, possessing a vast array of virulence factors ranging from adhesion and immunoevasion factors to siderophores and toxins (Liu, 2009). Because sortase enzymes are at the centre of a pathway which processes multiple virulence factors, targeting a sortase enzyme could effectively target a large number of virulence factors simultaneously. Also, as with many antibiotic targets, sortases are an exclusively bacterial family of enzymes, so a specific sortase inhibitor is unlikely to exhibit off-target effects. The closest functional homologues of sortase enzymes in humans are likely to be the structurally unrelated cysteine proteases such as the calpains, caspases and selected cathepsins.

For an organism to be a viable target for an anti-sortase therapy, several conditions must be met. The organism must possess a sortase, the sortase must anchor proteins which are essential for virulence, and inhibition of the sortase must abolish functional presentation of the essential proteins.

The discovery that the *S. aureus* sortase is essential for virulence in a mouse model (Mazmanian *et al.*, 2000) led to much interest in sortase inhibitors as therapeutics. The sole report of *in vivo* testing of a sortase inhibitor for treatment of *S. aureus* infection offers some encouraging results – administration of (Z)-3-(2,5-dimethoxyphenyl)-2-(4-methoxyphenyl) acrylonitrile (DMMA) reduced mortality in mouse models of kidney infection and joint infection, and while the authors noted some evidence of toxicity at high doses, the drug remained effective at a non-toxic dose. (Oh *et al.*, 2010).

It is disappointing that no further *in vivo* testing of sortase inhibitors has been reported, given the number of *S. aureus* sortase inhibitors which have been identified *in vitro*. It is not clear whether this is due to a genuine lack of experiments or a consequence of non-disclosure of results, positive or otherwise.

9.4 Contribution of This Work to the Field.

This work represents the first investigation of any kind into the sortase enzyme of *Clostridium difficile*, representing a valuable contribution to the fields of both sortase biology and the biology of the *C. difficile* cell surface. The major shortcomings of the work are the lack of a genetically complemented *C. difficile* 630 *srtB::erm* mutant and the reliance upon a C226A mutant *C. difficile* SrtB for structural studies. Detailed discussions on these points have been made in the relevant chapter discussions, along with plans for remedy of the shortcomings.

Adhesion is thought to be a key step for *C. difficile* infection, as histological investigations demonstrate clear association of bacteria with the mucosa (Goulding *et al.*, 2009). As seen in other toxigenic bacteria such as *C.diptheriae*, adhesion is likely to provide a twofold advantage to *C. difficile*, securing the bacterium's colonisation of its niche while ensuring that its toxins are delivered directly to the host. Although several investigations have identified proteins which appear to play a role in adhesion *in vitro* eg. Fbp68 (Hennequin *et al.*, 2003) and CbpA (Tulli *et al.*, 2013), no 'essential' *C. difficile* adhesin has yet been discovered. It is likely that in *C. difficile*, as in organisms such as *S. aureus*, adhesion is multifactorial. Multifactorial adhesion is an ideal case for sortase enzymes as targets of anti-infectives, because in the case of a species with multiple sortase-anchored adhesins, inactivating the sortase can prevent surface localisation of several adhesins simultaneously. The rationale of investigating the *C. difficile* sortase was to determine whether, as is clearly the case in *S. aureus*, the *C. difficile* sortase is required for virulence and therefore represents a novel target for a therapeutic.

At the time of conception of the work, it was a widely-held belief among many investigators that the sortase was an essential gene in *C. difficile* (Wren, 2011, Personal Communications *via* Dr Jon Kirby, Public Health England), based upon many unsuccessful attempts to inactivate the *srtB* gene using the Clostron system. However, such a hypothesis is not consistent with current knowledge of sortases in other species, as no sortase has yet been characterised as essential. It is also not consistent with the *in silico* predicted repertoire of the *C. difficile* sortase, as none of the predicted wall-sorted proteins possess putative functions of an essential nature. The results presented in Chapter 5 clearly contradict any suggestion that the *C. difficile* sortase is essential for survival of the organism – in fact, the rate of growth is identical in the wild type and sortase mutant.

The demonstration that the surface protein CD0386 is anchored to the peptidoglycan by SrtB provides a proof of principle that a sortase-mediated wall anchoring system operates in *C. difficile*. An opportunity now exists to comprehensively characterise the repertoire of the *C. difficile* sortase. The ideal approach would be comparative proteomic characterisation of the peptidoglycan-anchored proteome of the wild type and sortase knockout to definitively identify a population of

surface proteins, followed by biochemical, biophysical and structural investigation of those proteins identified as being anchored by the sortase.

The finding that the *C. difficile srtB::erm* mutant retains full virulence in the Syrian hamster model casts serious doubt upon the role of SrtB in pathogenesis. However, as discussed in Section 5.4.9, the hamster model is limited it its scope, and investigation of the sortase in an alternate model may be of merit if comprehensive studies on the phenotype of the mutant and *in vitro* characterisation of the sortase repertoire suggest that there are species-specific factors which may be poorly replicated in the hamster model.

Biochemical characterisation of SrtB, as described in Chapter 8, indicates a preference for an SPKTG sorting signal over an NVQTG signal, confirming the prediction of Comfort and Clubb (2004) and casting doubt upon the assertion of Tulli *et al* (2013) that the NVQTG-bearing collagen binding protein CbpA may be a sortase substrate. In addition, LC-MS analysis of the reaction products of SrtB and an SPKTG peptide clearly demonstrates that SrtB is a functional sortase capable of cleaving the peptide at the threonyl-glycyl bond and subsequently catalysing a transpeptidation with mDAP, the cross linking amino acid of the *C. difficile* peptidoglycan.

As discussed in Section 6.5.3, there are significant gaps in knowledge of the structure/function relationships of the sortase enzymes, particularly regarding recognition of their two substrates. The limited number of available sortase structures hampers interpretation, making it more difficult to identify structural elements which are conserved across all sortases or conserved within sortase classes. Once published, the *C. difficile* SrtB structure will be only the fourth SrtB structure and the third SrtB not associated with pilus biogenesis. By adding to the pool of SrtB structures, the *C. difficile* SrtB will aid in the general interpretation of sortases of all classes. As further experiments are performed in other species, re-interpretation of the structure of *C. difficile* SrtB may yield yet more information, and conversely further experiments regarding substrate recognition by the *C. difficile* SrtB will aid in interpretation of other sortases.

Early characterisations of the class B sortases of *S. aureus, B. anthracis and L. monocytogenes* led to the general classification of the class B subfamily as fulfilling a specialised role in iron regulation. The discovery that the *S. pyogenes* pilus-associated sortase is a class B enzyme provided the first indication that this generalisation may not be of value. Solving the structure of the *C. difficile* sortase has allowed its clear classification as a class B sortase, despite its putative repertoire lacking any association with iron acquisition or regulation. This casts clear doubt upon the paradigm. Indeed, the list of putative sortase-anchored proteins identified in *C. difficile* (see Section 5.3.2) is more suggestive of a ‘housekeeping’ sortase such as *S. aureus* SrtA.

While it is clear why bacteria such as *S. aureus* and *L. monocytogenes* have a regulatory necessity to segregate housekeeping sortase activity from specialised iron-associated anchoring, it is not clear why these enzymes are structurally different, since segregation is achieved by differences in the CWSS, and the regions of structural divergence (namely the additional helices) are not placed to directly interact with the substrates. Differences may be purely topological – the additional helix at the N-terminus of SrtB enzymes places the active site facing towards the cell surface, whereas in SrtA it faces away from the cell surface. Another hypothesis is that the differences in structure between SrtA and SrtB reflect wider differences in the cellular distribution or localisation of the enzymes. Studies into localisation of class A sortases in *S. mutans* and *S. pyogenes* (Hu *et al.*, 2008; Raz and Fischetti, 2008) have produced some very interesting results, and it would be very interesting to expand these investigations in other species and other sortase classes.

9.5 - Concluding Remarks

Within this thesis is described the first investigations into the sortase enzyme of the opportunistic pathogen *C. difficile*, comprised of *in vitro* and *in vivo* studies on a sortase gene knockout strain, structural characterisation of the sortase by X-ray crystallography, preliminary analysis of its reaction by biochemical techniques and crystallisation of one of the putative sortase-anchored proteins.

These results represent a novel contribution to the field of *C. difficile* cell biology and more generally add to the growing body of knowledge on the sortase enzyme family.

References

Aas, J., Gessert, C.E., Bakken, J.S., 2003. Recurrent *Clostridium difficile* colitis: case series involving 18 patients treated with donor stool administered via a nasogastric tube. *Clin. Infect. Dis.* 36, 580–585.

Akerlund, T., Persson, I., Unemo, M., Noren, T., Svenungsson, B., Wullt, M., Burman, L.G., 2008. Increased sporulation rate of epidemic *Clostridium difficile* Type 027/NAP1. *J. Clin. Microbiol.* 46, 1530–1533.

Alfa, M.J., Kabani, A., Lyerly, D., Moncrief, S., Neville, L.M., Al-Barrak, A., Harding, G.K.H., Dyck, B., Olekson, K., Embil, J.M., 2000. Characterization of a Toxin A-Negative, Toxin B-Positive Strain of *Clostridium difficile* Responsible for a Nosocomial Outbreak of *Clostridium difficile*-Associated Diarrhea. *J. Clin. Microbiol.* 38, 2706–2714.

Allen, S.J., Wareham, K., Bradley, C., Harris, W., Dhar, A., Brown, H., Foden, A., Cheung, W.Y., Gravenor, M.B., Plummer, S., Phillips, C.J., Mack, D., 2012. A multicentre randomised controlled trial evaluating lactobacilli and bifidobacteria in the prevention of antibiotic-associated diarrhoea in older people admitted to hospital: the PLACIDE study protocol. *BMC Infect. Dis.* 12, 108.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–10.

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201.

Aspevall, O., Lundberg, A., Burman, L.G., Akerlund, T., Svenungsson, B., 2006. Antimicrobial susceptibility pattern of *Clostridium difficile* and its relation to PCR ribotypes in a Swedish university hospital. *Antimicrob. Agents Chemother.* 50, 1890–1892.

Babcock, G.J., Broering, T.J., Hernandez, H.J., Mandell, R.B., Donahue, K., Boatright, N., Stack, A.M., Lowy, I., Graziano, R., Molrine, D., Ambrosino, D.M., Thomas, W.D., 2006. Human monoclonal antibodies directed against toxins A and B prevent *Clostridium difficile*-induced mortality in hamsters. *Infect. Immun.* 74, 6339–6347.

Baines, S.D., O'Connor, R., Freeman, J., Fawley, W.N., Harmanus, C., Mastrantonio, P., Kuijper, E.J., Wilcox, M.H., 2008. Emergence of reduced susceptibility to metronidazole in *Clostridium difficile*. *J. Antimicrob. Chemother.* 62, 1046–1052.

Bakker, D., Smits, W.K., Kuijper, E.J., Corver, J., 2012. TcdC does not significantly repress toxin expression in *Clostridium difficile* 630ΔErm. *PLoS One* 7, e43247.

Barbut, F., Decre, D., Lalande, V., Burghoffer, B., Noussair, L., Gigandon, A., Espinasse, F., Raskine, L., Robert, J., Mangeol, A., Branger, C., Petit, J.C., 2005. Clinical features of *Clostridium difficile*-associated diarrhoea due to binary toxin (actin-specific ADP-ribosyltransferase)-producing strains. *J. Med. Microbiol.* 54, 181–185.

Barbut, F., Gariazzo, B., Bonné, L., Lalande, V., Burghoffer, B., Luiuz, R., Petit, J.-C., 2007. Clinical features of *Clostridium difficile*-associated infections and molecular characterization

of strains: results of a retrospective study, 2000-2004. *Infect. Control Hosp. Epidemiol.* 28, 131–139.

Barnett, T.C., Scott, J.R., 2002. Differential recognition of surface proteins in *Streptococcus pyogenes* by two sortase gene homologs. *J. Bacteriol.* 184, 2181–2191.

Barocchi, M.A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Masignani, V., Hultenby, K., Taddei, A.R., Beiter, K., Wartha, F., von Euler, A., Covacci, A., Holden, D.W., Normark, S., Rappuoli, R., Henriques-Normark, B., 2006. A pneumococcal pilus influences virulence and host inflammatory responses. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2857–2862.

Barth, H., Blocker, D., Behlke, J., Bergsma-Schutter, W., Brisson, A., Benz, R., Aktories, K., 2000. Cellular uptake of Clostridium botulinum C2 toxin requires oligomerization and acidification. *J. Biol. Chem.* 275, 18704–18711.

Barth, H., Pfeifer, G., Hofmann, F., Maier, E., Benz, R., Aktories, K., 2001. Low pH-induced formation of ion channels by *Clostridium difficile* toxin B in target cells. *J. Biol. Chem.* 276, 10670–10676.

Bartlett, J.G., Onderdonk, A.B., Cisneros, R.L., Kasper, D.L., 1977. Clindamycin-associated colitis due to a toxin-producing species of Clostridium in hamsters. *J. Infect. Dis.* 136, 701–705.

Bentley, M.L., Gaweska, H., Kielec, J.M., McCafferty, D.G., 2007. Engineering the substrate specificity of *Staphylococcus aureus* Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J. Biol. Chem.* 282, 6571–81.

Bentley, M.L., Lamb, E.C., McCafferty, D.G., 2008. Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from *Staphylococcus aureus*. *J. Biol. Chem.* 283, 14762–14771.

Bernstein, E., Kaye, D., Abrutyn, E., Gross, P., Dorfman, M., Murasko, D.M., 1999. Immune response to influenza vaccination in a large healthy elderly population. *Vaccine* 17, 82–94.

Best, E.L., Freeman, J., Wilcox, M.H., 2012. Models for the study of *Clostridium difficile* infection. *Gut Microbes* 3, 145–167.

Bhangu, A., Nepogodiev, D., Gupta, A., Torrance, A., Singh, P., 2012. Systematic review and meta-analysis of outcomes following emergency surgery for *Clostridium difficile* colitis. *Br. J. Surg.* 99, 1501–1513.

Bierne, H., Garandeau, C., Pucciarelli, M.G., Sabet, C., Newton, S., Portillo, F.G., Cossart, P., Charbit, A., 2004. Sortase B , a New Class of Sortase in *Listeria monocytogenes*. *J. Bacteriol.* 186, 1972–1982.

Bignardi, G.E., 1998. Risk factors for *Clostridium difficile* infection. *J. Hosp. Infect.* 40, 1–15.

Bolton, R.P., Culshaw, M.A., 1986. Faecal metronidazole concentrations during oral and intravenous therapy for antibiotic associated colitis due to *Clostridium difficile*. *Gut* 27, 1169–1172.

Boneca, I.G., Dussurget, O., Cabanes, D., Nahori, M.-A., Sousa, S., Lecuit, M., Psylinakis, E., Bouriotis, V., Hugot, J.-P., Giovannini, M., Coyle, A., Bertin, J., Namane, A., Rousselle, J.-C., Cayet, N., Prévost, M.-C., Balloy, V., Chignard, M., Philpott, D.J., Cossart, P., Girardin, S.E., 2007. A critical role for peptidoglycan N-deacetylation in Listeria evasion from the host innate immune system. *Proc. Natl. Acad. Sci. U. S. A.* 104, 997–1002.

Brazier, J.S., 2001. Reduced susceptibility of*Clostridium difficile*to metronidazole. *J. Antimicrob. Chemother.* 48, 741–742.

Brazier, J.S., Raybould, R., Patel, B., Duckworth, G., Pearson, A., Charlett, A., Duerden, B.I., 2008. Distribution and antimicrobial susceptibility patterns of *Clostridium difficile* PCR ribotypes in English hospitals, 2007-08. *Euro Surveill. Eur. Commun. Dis. Bull.* 13.

Brünger, A.T., 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475.

Budzik, J.M., Marraffini, L.A., Schneewind, O., 2007. Assembly of pili on the surface of *Bacillus cereus* vegetative cells. *Mol. Microbiol.* 66, 495–510.

Budzik, J.M., Marraffini, L.A., Souda, P., Whitelegge, J.P., Faull, K.F., Schneewind, O., 2008a. Amide bonds assemble pili on the surface of bacilli. *Proc. Natl. Acad. Sci. U. S. A.* 105, 10215–10220.

Budzik, J.M., Oh, S.-Y., Schneewind, O., 2008b. Cell wall anchor structure of BcpA pili in *Bacillus anthracis*. *J. Biol. Chem.* 283, 36676–36686.

Budzik, J.M., Poor, C.B., Faull, K.F., Whitelegge, J.P., He, C., Schneewind, O., 2009. Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19992–19997.

Buffie, C.G., Jarchum, I., Equinda, M., Lipuma, L., Gobourne, A., Viale, A., Ubeda, C., Xavier, J., Pamer, E.G., 2012. Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to *Clostridium difficile*-induced colitis. *Infect. Immun.* 80, 62–73.

Burns, D.A., Heap, J.T., Minton, N.P., 2010. The diverse sporulation characteristics of *Clostridium difficile* clinical isolates are not associated with type. *Anaerobe* 16, 618–622.

Burns, D.A., Heeg, D., Cartman, S.T., Minton, N.P., 2011. Reconsidering the Sporulation Characteristics of Hypervirulent *Clostridium difficile* BI/NAP1/027. *PLoS One* 6, e24894.

Calabi, E., Phillips, A.D., Fairweather, N.F., 2002. Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues. *Infect. Immun.* 70, 5770–5778.

Calabi, E., Ward, S., Wren, B.W., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G., Fairweather, N.F., 2001. Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Mol. Microbiol.* 40, 1187–1199.

Carter, G.P., Douce, G.R., Govind, R., Howarth, P.M., Mackin, K.E., Spencer, J., Buckley, A.M., Antunes, A., Kotsanas, D., Jenkin, G. a., Dupuy, B., Rood, J.I., Lyras, D., 2011. The Anti-Sigma Factor TcdC Modulates Hypervirulence in an Epidemic BI/NAP1/027 Clinical Isolate

of *Clostridium difficile*. *PLoS Pathog.* 7, e1002317.

Cartman, S.T., Kelly, M.L., Heeg, D., Heap, J.T., Minton, N.P., 2012. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. *Appl. Environ. Microbiol.* 78, 4683–4690.

Chapetón Montes, D., Candela, T., Collignon, A., Janoir, C., 2011. Localization of the *Clostridium difficile* cysteine protease Cwp84 and insights into its maturation process. *J. Bacteriol.* 193, 5314–5321.

Chayen, N.E., 2005. Methods for separating nucleation and growth in protein crystallisation. *Prog. Biophys. Mol. Biol.* 88, 329–337.

Chen, X., Katchar, K., Goldsmith, J.D., Nanthakumar, N., Cheknis, A., Gerding, D.N., Kelly, C.P., 2008. A mouse model of *Clostridium difficile*-associated disease. *Gastroenterology* 135, 1984–1992.

Chen, Y., Zhang, W., Shi, Q., Hesek, D., Lee, M., Mobashery, S., Shoichet, B.K., 2009. Crystal structures of penicillin-binding protein 6 from *Escherichia coli*. *J. Am. Chem. Soc.* 131, 14345–14354.

Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., Farrow, J.A., 1994. The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. *Int. J. Syst. Bacteriol.* 44, 812–826.

Comfort, D., Clubb, R.T., 2004. A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. *Infect. Immun.* 72, 2710–2722.

Connolly, K.M., Smith, B.T., Pilpa, R., Ilangovan, U., Jung, M.E., Clubb, R.T., 2003. Sortase from *Staphylococcus aureus* does not contain a thiolate-imidazolium ion pair in its active site. *J. Biol. Chem.* 278, 34061–34065.

Courvalin, P., 2006. Vancomycin resistance in gram-positive cocci. *Clin. Infect. Dis.* 42 Suppl 1, S25–34.

Cousineau, B., Smith, D., Lawrence-Cavanagh, S., Mueller, J.E., Yang, J., Mills, D., Manias, D., Dunny, G., Lambowitz, A.M., Belfort, M., 1998. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* 94, 451–462.

Cunningham, R., Dale, B., Undy, B., Gaunt, N., 2003. Proton pump inhibitors as a risk factor for *Clostridium difficile* diarrhoea. *J. Hosp. Infect.* 54, 243–245.

Dallal, R.M., Harbrecht, B.G., Boujoukas, A.J., Sirio, C.A., Farkas, L.M., Lee, K.K., Simmons, R.L., 2002. Fulminant *Clostridium difficile*: an underappreciated and increasing cause of death and complications. *Ann. Surg.* 235, 363–372.

Danley, D.E., 2006. Crystallization to obtain protein-ligand complexes for structure-aided drug design. *Acta Crystallogr. D. Biol. Crystallogr.* 62, 569–75.

Davies, J.R., Svensater, G., Herzberg, M.C., 2009. Identification of novel LPXTG-linked surface

proteins from *Streptococcus gordonii*. *Microbiology* 155, 1977–1988.

Davies, N.L., Compson, J.E., Mackenzie, B., O'Dowd, V.L., Oxbrow, A.K.F., Heads, J.T., Turner, A., Sarkar, K., Dugdale, S.L., Jairaj, M., Christodoulou, L., Knight, D.E.O., Cross, A.S., Hervé, K.J.M., Tyson, K.L., Hailu, H., Doyle, C.B., Ellis, M., Kriek, M., Cox, M., Page, M.J.T., Moore, A.R., Lightwood, D.J., Humphreys, D.P., 2013. A mixture of functionally oligoclonal humanized monoclonal antibodies that neutralize *Clostridium difficile* TcdA and TcdB with high levels of in vitro potency shows in vivo protection in a hamster infection model. *Clin. Vaccine Immunol.* 20, 377–390.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., Richardson, D.C., 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375–W383.

De Jong, E., de Jong, A.S., Bartels, C.J.M., van der Rijt-van den Biggelaar, C., Melchers, W.J.G., Sturm, P.D.J., 2012. Clinical and laboratory evaluation of a real-time PCR for *Clostridium difficile* toxin A and B genes. *Eur. J. Clin. Microbiol. Infect. Dis.* 31, 2219–2225.

De la Riva, L., Willing, S.E., Tate, E.W., Fairweather, N.F., 2011. Roles of cysteine proteases Cwp84 and Cwp13 in biogenesis of the cell wall of *Clostridium difficile*. *J. Bacteriol.* 193, 3276–3285.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M., Gascuel, O., 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36, W465–9.

Dineen, S.S., McBride, S.M., Sonenshein, A.L., 2010. Integration of metabolism and virulence by *Clostridium difficile* CodY. *J. Bacteriol.* 192, 5350–5362.

Dramsi, S., Trieu-Cuot, P., Bierne, H., 2005. Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. *Res. Microbiol.* 156, 289–297.

Drudy, D., Fanning, S., Kyne, L., 2007a. Toxin A-negative, toxin B-positive *Clostridium difficile*. *Int. J. Infect. Dis.* 11, 5–10.

Drudy, D., Kyne, L., O'Mahony, R., Fanning, S., 2007b. gyrA mutations in fluoroquinolone-resistant *Clostridium difficile* PCR-027. *Emerg. Infect. Dis.* 13, 504–505.

Dryla, A., Gelbmann, D., von Gabain, A., Nagy, E., 2003. Identification of a novel iron regulated staphylococcal surface protein with haptoglobin-haemoglobin binding activity. *Mol. Microbiol.* 49, 37–53.

Egerer, M., Giesemann, T., Herrmann, C., Aktories, K., 2009. Autocatalytic processing of *Clostridium difficile* toxin B. Binding of inositol hexakisphosphate. *J. Biol. Chem.* 284, 3389–3395.

Egerer, M., Giesemann, T., Jank, T., Satchell, K.J.F., Aktories, K., 2007. Auto-catalytic cleavage of *Clostridium difficile* toxins A and B depends on cysteine protease activity. *J. Biol. Chem.* 282, 25314–25321.

El Mortaji, L., Fenel, D., Vernet, T., Di Guilmi, A.M., 2011. Association of RrgA and RrgC into the Streptococcus pneumoniae pilus by sortases C-2 and C-3. *Biochemistry* 51, 342–352.

Emerson, J.E., Reynolds, C.B., Fagan, R.P., Shaw, H. a, Goulding, D., Fairweather, N.F., 2009. A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. *Mol. Microbiol.* 74, 541–556.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 486–501.

Fagan, R.P., Albesa-Jové, D., Qazi, O., Svergun, D.I., Brown, K.A., Fairweather, N.F., 2009. Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. *Mol. Microbiol.* 71, 1308–1322.

Fagan, R.P., Fairweather, N.F., 2011. *Clostridium difficile* Has Two Parallel and Essential Sec Secretion Systems. *J. Biol. Chem.* 286, 27483–93.

Fan, J., Zhang, Y., Chuang-Smith, O.N., Frank, K.L., Guenther, B.D., Kern, M., Schlievert, P.M., Herzberg, M.C., 2012. Ecto-5'-nucleotidase: a candidate virulence factor in *Streptococcus sanguinis* experimental endocarditis. *PLoS One* 7, e38059.

Fenton, S., Stephenson, D., Weder, C., 1974. Pseudomembranous colitis associated with antibiotic therapy - an emerging entity. *Can. Med. Assoc. J.* 111, 1110–1114.

Flock, J.I., Fröman, G., Jönsson, K., Guss, B., Signäs, C., Nilsson, B., Raucci, G., Höök, M., Wadström, T., Lindberg, M., 1987. Cloning and expression of the gene for a fibronectin-binding protein from *Staphylococcus aureus*. *EMBO J.* 6, 2351–2357.

Fouet, A., Mesnage, S., Tosi-Couture, E., Gounon, P., Mock, M., 1999. *Bacillus anthracis* surface: capsule and S-layer. *J. Appl. Microbiol.* 87, 251–255.

Frankel, B.A., Kruger, R.G., Robinson, D.E., Kelleher, N.L., McCafferty, D.G., 2005. *Staphylococcus aureus* sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* 44, 11188–200.

Frankel, B.A., Tong, Y., Bentley, M.L., Fitzgerald, M.C., McCafferty, D.G., 2007a. Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. *Biochemistry* 46, 7269–7278.

Gadella, T.W.., 2009. FRET and FLIM Techniques. Elsevier Ltd.

Gao, X.W., Mubasher, M., Fang, C.Y., Reifer, C., Miller, L.E., 2010. Dose-response efficacy of a proprietary probiotic formula of *Lactobacillus acidophilu*s CL1285 and *Lactobacillus casei* LBC80R for antibiotic-associated diarrhea and *Clostridium difficile*-associated diarrhea prophylaxis in adult patients. *Am. J. Gastroenterol.* 105, 1636–1641.

Garandeau, C., Réglier-Poupet, H., Dubail, I., Beretti, J.-L., Berche, P., Charbit, A., 2002. The sortase SrtA of *Listeria monocytogenes* is involved in processing of internalin and in virulence. *Infect. Immun.* 70, 1382–90.

Gaspar, A.H., Marraffini, L.A., Elizabeth, M., Debord, K.L., Ton-that, H., Glass, E.M., 2005.

Bacillus anthracis Sortase A (SrtA) Anchors LPXTG Motif-Containing Surface Proteins to the Cell Wall Envelope *Bacillus anthracis* Sortase A (SrtA) Anchors LPXTG Motif-Containing Surface Proteins to the Cell Wall Envelope. *J. Bacteriol.* 187, 4646–4655.

Geerlof, A., Brown, J., Coutard, B., Egloff, M.P., Enguita, F.J., Fogg, M.J., Gilbert, R.J.C., Groves, M.R., Haouz, A., Nettleship, J.E., Nordlund, P., Owens, R.J., Ruff, M., Sainsbury, S., Svergun, D.I., Wilmanns, M., 2006. The impact of protein characterization in structural proteomics. *Acta Crystallogr. D. Biol. Crystallogr.* 62, 1125–1136.

George, R.H., Symonds, J.M., Dimock, F., Brown, J.D., Arabi, Y., Shinagawa, N., Keighley, M.R., Alexander-Williams, J., Burdon, D.W., 1978. Identification of *Clostridium difficile* as a cause of pseudomembranous colitis. *Br. Med. J.* 1, 695.

Gerding, D.N., 2004. Clindamycin, cephalosporins, fluoroquinolones, and *Clostridium difficile*-associated diarrhea: this is an antimicrobial resistance problem. *Clin. Infect. Dis.* 38, 646–648.

Geric, B., 2004. Distribution of *Clostridium difficile* variant toxinotypes and strains with binary toxin genes among clinical isolates in an American hospital. *J. Med. Microbiol.* 53, 887–894.

Giesemann, T., Jank, T., Gerhard, R., Maier, E., Just, I., Benz, R., Aktories, K., 2006. Cholesterol-dependent pore formation of *Clostridium difficile* toxin A. *J. Biol. Chem.* 281, 10808–10815.

Goldenberg, J.Z., Ma, S.S., Saxton, J.D., Martzen, M.R., Vandvik, P.O., Thorlund, K., Guyatt, G.H., Johnston, B.C., 1996. Cochrane Database of Systematic Reviews, *Cochrane database of systematic reviews (Online)*. John Wiley & Sons, Ltd, Chichester, UK.

Gonzales, M., Pepin, J., Frost, E.H., Carrier, J.C., Sirard, S., Fortier, L.-C., Valiquette, L., 2010. Faecal pharmacokinetics of orally administered vancomycin in patients with suspected *Clostridium difficile* infection. *BMC Infect. Dis.* 10, 363.

Goorhuis, A., Van der Kooi, T., Vaessen, N., Dekker, F.W., Van den Berg, R., Harmanus, C., van den Hof, S., Notermans, D.W., Kuijper, E.J., 2007. Spread and epidemiology of *Clostridium difficile* polymerase chain reaction ribotype 027/toxinotype III in The Netherlands. *Clin. Infect. Dis.* 45, 695–703.

Goulding, D., Thompson, H., Emerson, J., Fairweather, N.F., Dougan, G., Douce, G.R., 2009. Distinctive profiles of infection and pathology in hamsters infected with *Clostridium difficile* strains 630 and B1. *Infect. Immun.* 77, 5478–5485.

Govind, R., Dupuy, B., 2012. Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE. *PLoS Pathog.* 8, e1002727.

Greenberg, R.N., Marbury, T.C., Foglia, G., Warny, M., 2012. Phase I dose finding studies of an adjuvanted *Clostridium difficile* toxoid vaccine. *Vaccine* 30, 2245–2249.

Gülke, I., Pfeifer, G., Liese, J., Fritz, M., Hofmann, F., Aktories, K., Barth, H., 2001. Characterization of the enzymatic component of the ADP-ribosyltransferase toxin CDTa from *Clostridium difficile*. *Infect. Immun.* 69, 6004–6011.

Hall, I., O'Toole, E., 1935. Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, Bacillus difficilis. *Am. J. Dis. Child.* 49, 390.

Haynes, W., Lide, D., 2010. CRC handbook of chemistry and physics : a ready-reference book of chemical and physical data. Boca Raton, Fla.

Healthcare Comission, 2006. Investigation into outbreaks of *Clostridium difficile* at Stoke Mandeville Hospital, Buckinghamshire Hospitals NHS Trust.

Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P., Minton, N.P., 2007. The ClosTron: a universal gene knock-out system for the genus Clostridium. *J. Microbiol. Methods.* 70, 452–64.

Hempel, K., Herbst, F.-A., Moche, M., Hecker, M., Becher, D., 2011. Quantitative proteomic view on secreted, cell surface-associated, and cytoplasmic proteins of the methicillin-resistant human pathogen *Staphylococcus aureus* under iron-limited conditions. *J. Proteome Res.* 10, 1657–66.

Hendrickson, W.A., Horton, J.R., LeMaster, D.M., 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* 9, 1665–1672.

Hendrickson, W.A., Smith, J.L., Phizackerley, R.P., Merritt, E.A., 1988. Crystallographic structure analysis of lamprey hemoglobin from anomalous dispersion of synchrotron radiation. *Proteins* 4, 77–88.

Hennequin, C., Janoir, C., Barc, M.-C., Collignon, A., Karjalainen, T., 2003. Identification and characterization of a fibronectin-binding protein from *Clostridium difficile*. *Microbiology* 149, 2779–2787.

Heras, B., Martin, J.L., 2005. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr. D. Biol. Crystallogr.* 61, 1173–1180.

Hota, S.S., Achonu, C., Crowcroft, N.S., Harvey, B.J., Lauwers, A., Gardam, M.A., 2012. Determining mortality rates attributable to *Clostridium difficile* infection. *Emerg. Infect. Dis.* 18, 305–307.

Hu, P., Bian, Z., Fan, M., Huang, M., Zhang, P., 2008. Sec translocase and sortase A are colocalised in a locus in the cytoplasmic membrane of *Streptococcus mutans*. *Arch. Oral Biol.* 53, 150–154.

Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., Ellestad, G., 2003. Kinetic mechanism of *Staphylococcus aureus* sortase SrtA. *Biochemistry* 42, 11307–11315.

Hung, D.T., Shakhnovich, E.A., Pierson, E., Mekalanos, J.J., 2005. Small-molecule inhibitor of Vibrio cholerae virulence and intestinal colonization. *Science* 310, 670–674.

Hussain, H.A., 2005. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630 erm) and demonstration that the conjugative transposon Tn916 E enters the genome of this strain at multiple sites. *J. Med. Microbiol.* 54, 137–141.

ICM/BBC, 2008. Infections “the biggest NHS fear” [WWW Document]. URL http://news.bbc.co.uk/1/hi/health/7475561.stm (accessed 10.19.13).

Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., Clubb, R.T., 2001. Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6056–6061.

Izoré, T., Contreras-Martel, C., El Mortaji, L., Manzano, C., Terrasse, R., Vernet, T., Di Guilmi, A.M., Dessen, A., 2010. Structural basis of host cell recognition by the pilus adhesin from *Streptococcus pneumoniae*. *Structure* 18, 106–115.

Janarthanan, S., Ditah, I., Adler, D.G., Ehrinpreis, M.N., 2012. *Clostridium difficile*-associated diarrhea and proton pump inhibitor therapy: a meta-analysis. *Am. J. Gastroenterol.* 107, 1001–1010.

Jernberg, C., Löfmark, S., Edlund, C., Jansson, J.K., 2007. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *ISME J.* 1, 56–66.

Jonquières, R., Bierne, H., Fiedler, F., Gounon, P., Cossart, P., 1999. Interaction between the protein InlB of Listeria monocytogenes and lipoteichoic acid: a novel mechanism of protein association at the surface of gram-positive bacteria. *Mol. Microbiol.* 34, 902–14.

Jump, R.L.P., Pultz, M.J., Donskey, C.J., 2007. Vegetative *Clostridium difficile* survives in room air on moist surfaces and in gastric contents with reduced acidity: a potential mechanism to explain the association between proton pump inhibitors and C. difficile-associated diarrhea? *Antimicrob. Agents Chemother.* 51, 2883–2887.

Just, I., Selzer, J., Wilm, M., von Eichel-Streiber, C., Mann, M., Aktories, K., 1995a. Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* 375, 500–503.

Just, I., Wilm, M., Selzer, J., Rex, G., von Eichel-Streiber, C., Mann, M., Aktories, K., 1995b. The enterotoxin from *Clostridium difficile* (ToxA) monoglucosylates the Rho proteins. *J. Biol. Chem.* 270, 13932–13936.

Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.

Kang, H.J., Coulibaly, F., Proft, T., Baker, E.N., 2011. Crystal structure of Spy0129, a *Streptococcus pyogenes* class B sortase involved in pilus assembly. *PLoS One* 6, e15969.

Kang, H.J., Paterson, N.G., Gaspar, A.H., Ton-That, H., Baker, E.N., 2009. The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc. Natl. Acad. Sci. U. S. A.* 106, 16967–16971.

Kassam, Z., Lee, C.H., Yuan, Y., Hunt, R.H., 2013. Fecal microbiota transplantation for *Clostridium difficile* infection: systematic review and meta-analysis. *Am. J. Gastroenterol.* 108, 500–508.

Kelley, L.A., Sternberg, M.J.E., 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* 4, 363–371.

Khanna, S., Pardi, D.S., Aronson, S.L., Kammer, P.P., Orenstein, R., St Sauver, J.L., Harmsen, W.S., Zinsmeister, A.R., 2012. The epidemiology of community-acquired *Clostridium difficile* infection: a population-based study. *Am. J. Gastroenterol.* 107, 89–95.

Khare, B., Krishnan, V., Rajashankar, K.R., Xin, M., Narayana, S. V, 2011. Structural Differences between the *Streptococcus agalactiae* Housekeeping and Pilus-Specific Sortases : SrtA and SrtC1. *PLoS One* 6, e22995.

Kim, H.K., Emolo, C., DeDent, A.C., Falugi, F., Missiakas, D.M., Schneewind, O., 2012. Protein A-specific monoclonal antibodies and prevention of *Staphylococcus aureus* disease in mice. *Infect. Immun.* 80, 3460–3470.

Kirby, J.M., 2011. The Pathogenesis of *Clostridium difficile* infection. University of Bath.

Kirby, J.M., Ahern, H., Roberts, A.K., Kumar, V., Freeman, Z., Acharya, K.R., Shone, C.C., 2009. Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of *Clostridium difficile*. *J. Biol. Chem.* 284, 34666–34673.

Kline, K. a, Kau, A.L., Chen, S.L., Lim, A., Pinkner, J.S., Rosch, J., Nallapareddy, S.R., Murray, B.E., Henriques-Normark, B., Beatty, W., Caparon, M.G., Hultgren, S.J., 2009. Mechanism for sortase localization and the role of sortase localization in efficient pilus assembly in *Enterococcus faecalis*. *J. Bacteriol.* 191, 3237–3247.

Kotila, S.M., Virolainen, A., Snellman, M., Ibrahem, S., Jalava, J., Lyytikäinen, O., 2011. Incidence, case fatality and genotypes causing *Clostridium difficile* infections, Finland, 2008. *Clin. Microbiol. Infect.* 17, 888–893.

Krishnan, V., Gaspar, A.H., Ye, N., Mandlik, A., Ton-That, H., Narayana, S.V.L., 2007. An IgG-like domain in the minor pilin GBS52 of Streptococcus agalactiae mediates lung epithelial cell adhesion. *Structure* 15, 893–903.

Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D. Biol. Crystallogr.* 60, 2256–2268.

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797.

Kruger, R.G., Barkallah, S., Frankel, B.A., McCafferty, D.G., 2004a. Inhibition of the *Staphylococcus aureus* sortase transpeptidase SrtA by phosphinic peptidomimetics. *Bioorg. Med. Chem.* 12, 3723–9.

Kruger, R.G., Dostal, P., McCafferty, D.G., 2004b. Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the *Staphylococcus aureus* sortase transpeptidase SrtA. *Anal. Biochem.* 326, 42–8.

Kruger, R.G., Dostal, P., McCafferty, D.G., 2004c. Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the *Staphylococcus aureus* sortase transpeptidase SrtA. *Anal. Biochem.* 326, 42–8.

Kruger, R.G., Otvos, B., Frankel, B.A., Bentley, M.L., Dostal, P., McCafferty, D.G., 2004d. Analysis of the substrate specificity of the *Staphylococcus aureus* sortase transpeptidase SrtA. *Biochemistry* 43, 1541–1551.

Kutty, P.K., Woods, C.W., Sena, A.C., Benoit, S.R., Naggie, S., Frederick, J., Evans, S., Engel, J.,

McDonald, L.C., 2010. Risk factors for and estimated incidence of community-associated *Clostridium difficile* infection, North Carolina, USA. *Emerg. Infect. Dis.* 16, 197–204.

Kvach, E.J., Ferguson, D., Riska, P.F., Landry, M.L., 2010. Comparison of BD GeneOhm Cdiff real-time PCR assay with a two-step algorithm and a toxin A/B enzyme-linked immunosorbent assay for diagnosis of toxigenic *Clostridium difficile* infection. *J. Clin. Microbiol.* 48, 109–114.

Kwok, C.S., Arthur, A.K., Anibueze, C.I., Singh, S., Cavallazzi, R., Loke, Y.K., 2012. Risk of *Clostridium difficile* infection with acid suppressing drugs and antibiotics: meta-analysis. *Am. J. Gastroenterol.* 107, 1011–1019.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–8.

Larson, H.E., Barclay, F.E., Honour, P., Hill, I.D., 1982. Epidemiology of *Clostridium difficile* in Infants. *J. Infect. Dis.* 146, 727–733.

Lawley, T.D., Clare, S., Walker, A.W., Goulding, D., Stabler, R.A., Croucher, N., Mastroeni, P., Scott, P., Raisen, C., Mottram, L., Fairweather, N.F., Wren, B.W., Parkhill, J., Dougan, G., 2009. Antibiotic treatment of *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. *Infect. Immun.* 77, 3661–3669.

Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N., Breaker, R.R., 2010. An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger. *Science* 329, 845–848.

Lewis, S.D., Johnson, F.A., Shafer, J.A., 1981. Effect of cysteine-25 on the ionization of histidine-159 in papain as determined by proton nuclear magnetic resonance spectroscopy. Evidence for a his-159--Cys-25 ion pair and its possible role in catalysis. *Biochemistry* 20, 48–51.

Liew, C.K., Smith, B.T., Pilpa, R., Suree, N., Ilangovan, U., Connolly, K.M., Jung, M.E., Clubb, R.T., 2004. Localization and mutagenesis of the sorting signal binding site on sortase A from *Staphylococcus aureus*. *FEBS Lett.* 571, 221–226.

Liew, P.X., Wang, C.L.C., Wong, S.-L., 2011. Functional characterization of a *Bacillus subtilis* sortase and its substrate and its use to covalently anchor a heterologous protein to the B. subtilis cell wall for surface display. *J. Bacteriol.*

Liu, G.Y., 2009. Molecular pathogenesis of *Staphylococcus aureus* infection. *Pediatr. Res.* 65, 71R–77R.

Loo, V.G., Bourgault, A.-M., Poirier, L., Lamothe, F., Michaud, S., Turgeon, N., Toye, B., Beaudoin, A., Frost, E.H., Gilca, R., Brassard, P., Dendukuri, N., Béliveau, C., Oughton, M., Brukner, I., Dascal, A., 2011. Host and pathogen factors for *Clostridium difficile* infection and colonization. *N. Engl. J. Med.* 365, 1693–1703.

Louie, T.J., Cannon, K., Byrne, B., Emery, J., Ward, L., Eyben, M., Krulicki, W., 2012. Fidaxomicin preserves the intestinal microbiome during and after treatment of *Clostridium difficile* infection (CDI) and reduces both toxin reexpression and recurrence of CDI. *Clin.*

Infect. Dis. 55 Suppl 2, S132–142.

Louie, T.J., Miller, M.A., Mullane, K.M., Weiss, K., Lentnek, A., Golan, Y., Gorbach, S., Sears, P., Shue, Y.-K., 2011. Fidaxomicin versus vancomycin for *Clostridium difficile* infection. *N. Engl. J. Med.* 364, 422–431.

Lyras, D., O'Connor, J.R., Howarth, P.M., Sambol, S.P., Carter, G.P., Phumoonna, T., Poon, R., Adams, V., Vedantam, G., Johnson, S., Gerding, D.N., Rood, J.I., 2009. Toxin B is essential for virulence of *Clostridium difficile*. *Nature* 458, 1176–1179.

MacCannell, D.R., Louie, T.J., Gregson, D.B., Laverdiere, M., Labbe, A.-C., Laing, F., Henwick, S., 2006. Molecular analysis of *Clostridium difficile* PCR ribotype 027 isolates from Eastern and Western Canada. *J. Clin. Microbiol.* 44, 2147–2152.

Mangel, W.F., Singer, P.T., Cyr, D.M., Umland, T.C., Toledo, D.L., Stroud, R.M., Pflugrath, J.W., Sweet, R.M., 1990. Structure of an acyl-enzyme intermediate during catalysis: (guanidinobenzoyl)trypsin. *Biochemistry* 29, 8351–8357.

Manzano, C., Contreras-Martel, C., El Mortaji, L., Izoré, T., Fenel, D., Vernet, T., Schoehn, G., Di Guilmi, A.M., Dessen, A., 2008. Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*. *Structure* 16, 1838–48.

Marchler-Bauer, A., Bryant, S.H., 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 32, W327–31.

Maresso, A.W., Chapa, T.J., Schneewind, O., 2006. Surface protein IsdC and Sortase B are required for heme-iron scavenging of *Bacillus anthracis*. *J. Bacteriol.* 188, 8145–8152.

Maresso, A.W., Wu, R., Kern, J.W., Zhang, R., Janik, D., Missiakas, D.M., Duban, M.-E., Joachimiak, A., Schneewind, O., 2007. Activation of inhibitors by sortase triggers irreversible modification of the active site. *J. Biol. Chem.* 282, 23129–23139.

Mariscotti, J.F., Garcia-del Portillo, F., Pucciarelli, M.G., 2009. The Listeria monocytogenes sortase-B recognizes varied amino acids at position 2 of the sorting motif. *J. Biol. Chem.* 284, 6140–6146.

Marraffini, L.A., Schneewind, O., 2006. Targeting proteins to the cell wall of sporulating *Bacillus anthracis*. *Mol. Microbiol.* 62, 1402–1417.

Marraffini, L.A., Schneewind, O., 2007. Sortase C-mediated anchoring of BasI to the cell wall envelope of *Bacillus anthracis*. *J. Bacteriol.* 189, 6425–6436.

Masuda, K., Itoh, M., Kawata, T., 1989. Characterization and reassembly of a regular array in the cell wall of *Clostridium difficile* GAI 4131. *Microbiol. Immunol.* 33, 287–98.

Matamouros, S., England, P., Dupuy, B., 2007. *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Mol. Microbiol.* 64, 1274–1288.

Matsuki, S., Ozaki, E., Shozu, M., Inoue, M., Shimizu, S., Yamaguchi, N., Karasawa, T., Yamagishi, T., Nakamura, S., 2005. Colonization by *Clostridium difficile* of neonates in a hospital, and infants and children in three day-care facilities of Kanazawa, Japan. *Int.*

Microbiol. 8, 43–8.

Mazmanian, S.K., Liu, G., Jensen, E.R., Lenoy, E., Schneewind, O., 2000. *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proc. Natl. Acad. Sci. U. S. A.* 4220, 5510–5515.

Mazmanian, S.K., Liu, G., Ton-That, H., Schneewind, O., 1999. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* 285, 760–763.

Mazmanian, S.K., Ton-That, H., Su, K., Schneewind, O., 2002. An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 2293–2298.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.

McCusker, M.E., Harris, A.D., Perencevich, E., Roghmann, M.-C., 2003. Fluoroquinolone use and *Clostridium difficile*-associated diarrhea. *Emerg. Infect. Dis.* 9, 730–733.

McDevitt, D., Francois, P., Vaudaux, P., Foster, T.J., 1994. Molecular characterization of the clumping factor (fibrinogen receptor) of *Staphylococcus aureus*. *Mol. Microbiol.* 11, 237–248.

McDonough, M.A., Anderson, J.W., Silvaggi, N.R., Pratt, R.F., Knox, J.R., Kelly, J.A., 2002. Structures of two kinetic intermediates reveal species specificity of penicillin-binding proteins. *J. Mol. Biol.* 322, 111–122.

Merrigan, M., Venugopal, A., Mallozzi, M., Roxas, B., Viswanathan, V.K., Johnson, S., Gerding, D.N., Vedantam, G., 2010. Human Hypervirulent *Clostridium difficile* Strains Exhibit Increased Sporulation as well as Robust Toxin Production. *J. Bacteriol.* 192, 4904–4911.

Migliori, G., 2007. First tuberculosis cases in Italy resistant to all tested drugs. *Euro Surveill. Eur. Commun. Dis. Bull.* 12.

Miller, B., Wheeler, M., 1975. Letter: Clindamycin-associated pseudomembranous colitis. *Br. Med. J.* 3, 433–434.

Miyajima, F., Roberts, P., Swale, A., Price, V., Jones, M., Horan, M., Beeching, N., Brazier, J.S., Parry, C., Pendleton, N., Pirmohamed, M., 2011. Characterisation and carriage ratio of *Clostridium difficile* strains isolated from a community-dwelling elderly population in the United Kingdom. *PLoS One* 6, e22804.

Mohr, G., Smith, D., Belfort, M., Lambowitz, A.M., 2000. Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev.* 14, 559–73.

Moncrief, J.S., Barroso, L.A., Wilkins, T.D., 1997. Positive regulation of *Clostridium difficile* toxins. *Infect. Immun.* 65, 1105–1108.

Morgan, O.W., Rodrigues, B., Elston, T., Verlander, N.Q., Brown, D.F.J., Brazier, J.S., Reacher, M., 2008. Clinical severity of *Clostridium difficile* PCR ribotype 027: a case-case study. *PLoS*

One 3, e1812.

Morris, A.M., Jobe, B.A., Stoney, M., Sheppard, B.C., Deveney, C.W., Deveney, K.E., 2002. *Clostridium difficile* colitis: an increasingly aggressive iatrogenic disease? *Arch. Surg.* 137, 1096–1100.

Mullany, P., Wilks, M., Puckey, L., Tabaqchali, S., 1994. Gene cloning in *Clostridium difficile* using Tn916 as a shuttle conjugative transposon. *Plasmid* 31, 320–3.

Mullany, P., Wilks, M., Tabaqchali, S., 1991. Transfer of Tn916 and Tn916 delta E into *Clostridium difficile*: demonstration of a hot-spot for these elements in the C. difficile genome. *FEMS Microbiol. Lett.* 63, 191–4.

Mullany, P., Williams, R., Langridge, G.C., Turner, D.J., Whalan, R., Clayton, C., Lawley, T.D., Hussain, H.A., McCurrie, K., Morden, N., Allan, E., Roberts, A.P., 2012. Behavior and target site selection of conjugative transposon Tn916 in two different strains of toxigenic *Clostridium difficile*. *Appl. Environ. Microbiol.* 78, 2147–53.

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* 53, 240–255.

Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540.

Na, X., Kim, H., Moyer, M.P., Pothoulakis, C., LaMont, J.T., 2008. gp96 is a human colonocyte plasma membrane binding protein for *Clostridium difficile* toxin A. *Infect. Immun.* 76, 2862–2871.

Naik, M.T., Suree, N., Ilangovan, U., Liew, C.K., Thieu, W., Campbell, D.O., Clemens, J.J., Jung, M.E., Clubb, R.T., 2006. *Staphylococcus aureus* Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. *J. Biol. Chem.* 281, 1817–1826.

Navarre, W.W., 1998. Anchor Structure of Staphylococcal Surface Proteins. II. COOH-terminal structure of muramidase and amidase-solubilized surface protein. *J. Biol. Chem.* 273, 29135–29142.

Navarre, W.W., Schneewind, O., 1994. Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. *Mol. Microbiol.* 14, 115–121.

Necchi, F., Nardi-Dei, V., Biagini, M., Assfalg, M., Nuccitelli, A., Cozzi, R., Norais, N., Telford, J.L., Rinaudo, C.D., Grandi, G., Maione, D., 2011. Sortase A Substrate Specificity in GBS Pilus 2a Cell Wall Anchoring. *PLoS One* 6, e25300.

Nerandzic, M.M., Pultz, M.J., Donskey, C.J., 2009. Examination of potential mechanisms to explain the association between proton pump inhibitors and *Clostridium difficile* infection. *Antimicrob. Agents Chemother.* 53, 4133–4137.

Newton, S.M.C., Klebba, P.E., Raynaud, C., Shao, Y., Jiang, X., Dubail, I., Archer, C., Frehel, C., Charbit, A., 2005. The svpA-srtB locus of *Listeria monocytogenes*: fur-mediated iron

regulation and effect on virulence. *Mol. Microbiol.* 55, 927–940.

Nobbs, A.H., Vajna, R.M., Johnson, J.R., Zhang, Y., Erlandsen, S.L., Oli, M.W., Kreth, J., Brady, L.J., Herzberg, M.C., 2007. Consequences of a sortase A mutation in *Streptococcus gordonii*. *Microbiology* 153, 4088–97.

Oh, K.-B., Nam, K.-W., Ahn, H., Shin, J., Kim, S., Mar, W., 2010. Therapeutic effect of (Z)-3-(2,5-dimethoxyphenyl)-2-(4-methoxyphenyl) acrylonitrile (DMMA) against *Staphylococcus aureus* infection in a murine model. *Biochem. Biophys. Res. Commun.* 396, 440–444.

Oke, M., Carter, L.G., Johnson, K.A., Liu, H., McMahon, S.A., Yan, X., Kerou, M., Weikart, N.D., Kadi, N., Sheikh, M.A., Schmelz, S., Dorward, M., Zawadzki, M., Cozens, C., Falconer, H., Powers, H., Overton, I.M., van Niekerk, C.A.J., Peng, X., Patel, P., Garrett, R.A., Prangishvili, D., Botting, C.H., Coote, P.J., Dryden, D.T.F., Barton, G.J., Schwarz-Linek, U., Challis, G.L., Taylor, G.L., White, M.F., Naismith, J.H., 2010. The Scottish Structural Proteomics Facility: targets, methods and outputs. *J. Struct. Funct. Genomics* 11, 167–180.

Olling, A., Seehase, S., Minton, N.P., Tatge, H., Schröter, S., Kohlscheen, S., Pich, A., Just, I., Gerhard, R., 2012. Release of TcdA and TcdB from *Clostridium difficile* cdi 630 is not affected by functional inactivation of the tcdE gene. *Microb. Pathog.* 52, 92–100.

Pallen, M.J., Lam, A.C., Antonio, M., Dunbar, K., 2001. An embarrassment of sortases - a richness of substrates? *Trends Microbiol.* 9, 97–102.

Palmqvist, N., Foster, T., Tarkowski, A., Josefsson, E., 2002. Protein A is a virulence factor in *Staphylococcus aureus* arthritis and septic death. *Microb. Pathog.* 33, 239–249.

Pantosti, A., Cerquetti, M., Viti, F., Ortisi, G., Mastrantonio, P., 1989. Immunoblot analysis of serum immunoglobulin G response to surface proteins of *Clostridium difficile* in patients with antibiotic-associated diarrhea. *J. Clin. Microbiol* 27, 2594–2597.

Papatheodorou, P., Carette, J.E., Bell, G.W., Schwan, C., Guttenberg, G., Brummelkamp, T.R., Aktories, K., 2011. Lipolysis-stimulated lipoprotein receptor (LSR) is the host receptor for the binary toxin *Clostridium difficile* transferase (CDT). *Proc. Natl. Acad. Sci. U. S. A.* 108, 16422–16427.

Papatheodorou, P., Zamboglou, C., Genisyuerek, S., Guttenberg, G., Aktories, K., 2010. Clostridial glucosylating toxins enter cells via clathrin-mediated endocytosis. *PLoS One* 5, e10673.

Paterson, N.G., Baker, E.N., 2011. Structure of the Full-Length Major Pilin from *Streptococcus pneumoniae:* Implications for Isopeptide Bond Formation in Gram-Positive Bacterial Pili. *PLoS One* 6, e22095.

Peltier, J., Courtin, P., El Meouche, I., Lemée, L., Chapot-Chartier, M.-P., Pons, J.-L., 2011. *Clostridium difficile* has an original peptidoglycan structure with a high level of N-acetylglucosamine deacetylation and mainly 3-3 cross-links. *J. Biol. Chem.* 286, 29053–29062.

Pepin, J., Saheb, N., Coulombe, M.-A., Alary, M.-E., Corriveau, M.-P., Authier, S., Leblanc, M., Rivard, G., Bettez, M., Primeau, V., Nguyen, M., Jacob, C.-E., Lanthier, L., 2005a. Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*-

associated diarrhea: a cohort study during an epidemic in Quebec. *Clin. Infect. Dis.* 41, 1254–1260.

Pepin, J., Valiquette, L., Alary, M.-E., Villemure, P., Pelletier, A., Forget, K., Pépin, K., Chouinard, D., 2004. *Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. *CMAJ* 171, 466–472.

Pepin, J., Valiquette, L., Cossette, B., 2005b. Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *CMAJ* 173, 1037–1042.

Perelle, S., Gibert, M., Bourlioux, P., Corthier, G., Popoff, M.R., 1997. Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infect. Immun.* 65, 1402–1407.

Persson, K., 2011. Structure of the sortase AcSrtC-1 from *Actinomyces oris. Acta Crystallogr. D. Biol. Crystallogr.* 67, 212–217.

Persson, K., Esberg, A., Claesson, R., Strömberg, N., 2012. The pilin protein FimP from Actinomyces oris: crystal structure and sequence analyses. *PLoS One* 7, e48364.

Perutz, M.F., Rossman, M.G., Cullis, A.F., Muirhead, H., Will, G., North, A.C.T., 1960. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* 185, 416–422.

Peterfreund, G.L., Vandivier, L.E., Sinha, R., Marozsan, A.J., Olson, W.C., Zhu, J., Bushman, F.D., 2012. Succession in the gut microbiome following antibiotic and antibody therapies for *Clostridium difficile*. *PLoS One* 7, e46966.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–6.

Pfeifer, G., Schirmer, J., Leemhuis, J., Busch, C., Meyer, D.K., Aktories, K., Barth, H., 2003. Cellular uptake of *Clostridium difficile* toxin B. Translocation of the N-terminal catalytic domain into the cytosol of eukaryotic cells. *J. Biol. Chem.* 278, 44535–44341.

Pinitglang, S., Watts, A.B., Patel, M., Reid, J.D., Noble, M.A., Gul, S., Bokth, A., Naeem, A., Patel, H., Thomas, E.W., Sreedharan, S.K., Verma, C., Brocklehurst, K., 1997. A classical enzyme active center motif lacks catalytic competence until modulated electrostatically. *Biochemistry* 36, 9968–9982.

Piotukh, K., Geltinger, B., Heinrich, N., Gerth, F., Beyermann, M., Freund, C., Schwarzer, D., 2011. Directed Evolution of Sortase A Mutants with Altered Substrate Selectivity Profiles. *J. Am. Chem. Soc.* 17–20.

Pointon, J.A., Smith, W.D., Saalbach, G., Crow, A., Kehoe, M.A., Banfield, M.J., 2010. A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction. *J. Biol. Chem.* 285, 33858–33866.

Popoff, M.R., Rubin, E.J., Gill, D.M., Boquet, P., 1988. Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infect. Immun.* 56, 2299–2306.

Powers, J.H., 2004. Antimicrobial drug development--the past, the present, and the future. *Clin. Microbiol. Infect.* 10 Suppl 4, 23–31.

Pucciarelli, M.G., Calvo, E., Sabet, C., Bierne, H., Cossart, P., García-del Portillo, F., 2005. Identification of substrates of the *Listeria monocytogenes* sortases A and B by a non-gel proteomic analysis. *Proteomics* 5, 4808–4817.

Purdy, D., O'Keeffe, T. a T., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., Minton, N.P., 2002. Conjugative transfer of clostridial shuttle vectors from Escherichia coli to *Clostridium difficile* through circumvention of the restriction barrier. *Mol. Microbiol.* 46, 439–52.

Race, P.R., Bentley, M.L., Melvin, J.A., Crow, A., Hughes, R.K., Smith, W.D., Sessions, R.B., Kehoe, M.A., McCafferty, D.G., Banfield, M.J., 2009. Crystal structure of *Streptococcus pyogenes* sortase A: implications for sortase mechanism. *J. Biol. Chem.* 284, 6924–6933.

Ramachandran, G.N., Sasisekharan, V., 1968. Conformation of polypeptides and proteins. *Adv. Protein Chem.* 23, 283–438.

Rampelli, S., Candela, M., Severgnini, M., Biagi, E., Turroni, S., Roselli, M., Carnevali, P., Donini, L., Brigidi, P., 2013. A probiotics-containing biscuit modulates the intestinal microbiota in the elderly. *J. Nutr. Health Aging* 17, 166–172.

Raz, A., Fischetti, V.A., 2008. Sortase A localizes to distinct foci on the Streptococcus pyogenes membrane. *Proc. Natl. Acad. Sci. U. S. A* 2008, 1–6.

Reichert, J.M., 2013. Which are the antibodies to watch in 2013? *MAbs* 5, 1–4.

Reynolds, C.B., Emerson, J.E., de la Riva, L., Fagan, R.P., Fairweather, N.F., 2011. The *Clostridium difficile* cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. *PLoS Pathog.* 7, e1002024.

Rifkin, G.D., Fekety, F.R., Silva, J., 1977. Antibiotic-induced colitis implication of a toxin neutralised by *Clostridium sordellii* antitoxin. *Lancet* 2, 1103–1106.

Riggs, M.M., Sethi, A.K., Zabarsky, T.F., Eckstein, E.C., Jump, R.L.P., Donskey, C.J., 2007. Asymptomatic Carriers Are a Potential Source for Transmission of Epidemic and Nonepidemic *Clostridium difficile* Strains among Long-Term Care Facility Residents. *Clin. Infect. Dis.* 45, 992–998.

Roberts, A., McGlashan, J., Al-Abdulla, I., Ling, R., Denton, H., Green, S., Coxon, R., Landon, J., Shone, C., 2012. Development and evaluation of an ovine antibody-based platform for treatment of *Clostridium difficile* infection. *Infect. Immun.* 80, 875–882.

Robson, S.A., Jacobitz, A.W., Phillips, M.L., Clubb, R.T., 2012. Solution structure of the sortase required for efficient production of infectious *Bacillus anthracis* spores. *Biochemistry* 51, 7953–7963.

Rohlke, F., Surawicz, C.M., Stollman, N., 2010. Fecal flora reconstitution for recurrent *Clostridium difficile* infection: results and methodology. *J. Clin. Gastroenterol.* 44, 567–570.

Rossman, M.G., Blow, D.M., 1962. The Detection of Sub-Units within the Crystallographic Asymmetric Unit. *Acta Cryst.* 15, 24–32.

Rousseau, C., Poilane, I., De Pontual, L., Maherault, A.-C., Le Monnier, A., Collignon, A., 2012. *Clostridium difficile* Carriage in Healthy Infants in the Community: A Potential Reservoir for Pathogenic Strains. *Clin. Infect. Dis.* 55, 1209–1215.

Rupnik, M., Wilcox, M.H., Gerding, D.N., 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat. Rev. Microbiol.* 7, 526–536.

Scaria, J., Janvilisri, T., Fubini, S., Gleed, R.D., Mcdonough, S.P., Chang, Y., 2011. *Clostridium difficile* Transcriptome Analysis Using Pig Ligated Loop Model Reveals Modulation of Pathways Not Modulated In Vitro *J. Infect. Dis*. 203, 1613–1620.

Schleifer, K.H., Kandler, O., 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* 36, 407–477.

Schneewind, O., Model, P., Fischetti, V.A., 1992. Sorting of protein A to the staphylococcal cell wall. *Cell* 70, 267–281.

Schoster, A., Kokotovic, B., Permin, A., Pedersen, P.D., Dal Bello, F., Guardabassi, L., 2013. In vitro inhibition of *Clostridium difficile* and Clostridium perfringens by commercial probiotic strains. *Anaerobe* 20, 36–41.

Schwan, C., Stecher, B., Tzivelekidis, T., van Ham, M., Rohde, M., Hardt, W.-D., Wehland, J., Aktories, K., 2009. *Clostridium difficile* toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria. *PLoS Pathog.* 5, e1000626.

Scott, C.J., Dowell, A.M.C., Martin, S.L., Lynas, J.F., Vandenbroeck, K., Walker, B., 2002. Irreversible inhibition of the bacterial cysteine protease-transpeptidase sortase (SrtA) by substrate-derived affinity labels *Biochem. J.* 958, 953–958.

Selinummi, J., Seppälä, J., Yli-Harja, O., Puhakka, J.A., 2005. Software for quantification of labeled bacteria from digital microscope images by automated image analysis. *Biotechniques* 39, 859–63.

Shahravan, H., Qu, X., Chan, Is., Shin, J., 2008. Enhancing the specificity of the enterokinase cleavage reaction to promote efficient cleavage of a fusion tag. *Protein Expr. Purif.* 59, 314–319.

Shakhnovich, E.A., Hung, D.T., Pierson, E., Lee, K., Mekalanos, J.J., 2007. Virstatin inhibits dimerization of the transcriptional activator ToxT. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2372–2377.

Shivers, R.P., Sonenshein, A.L., 2004. Activation of the Bacillus subtilis global regulator CodY by direct interaction with branched-chain amino acids. *Mol. Microbiol.* 53, 599–611.

Sirard, S., Valiquette, L., Fortier, L.-C., 2011. Lack of association between clinical outcome of *Clostridium difficile* infections, strain type, and virulence-associated phenotypes. *J. Clin. Microbiol.* 49, 4060–4066.

Smith, D., Zhong, J., Matsuura, M., Lambowitz, A.M., Belfort, M., 2005. Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev.* 19, 2477–87.

Smith, King, 1962. Occurrence of *Clostridium difficile* in infections of man. *J. Bacteriol.* 84, 65–7.

Spraggon, G., Koesema, E., Scarselli, M., Malito, E., Biagini, M., Norais, N., Emolo, C., Barocchi, M.A., Giusti, F., Hilleringmann, M., Rappuoli, R., Lesley, S., Covacci, A., Masignani, V., Ferlenghi, I., 2010. Supramolecular organization of the repetitive backbone unit of the Streptococcus pneumoniae pilus. *PLoS One* 5, e10919.

Stabler, R.A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T.D., Sebaihia, M., Quail, M.A., Rose, G., Gerding, D.N., Gibert, M., Popoff, M.R., Parkhill, J., Dougan, G., Wren, B.W., 2009. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol* 10, R102.

Stark, P.L., Lee, A., Parsonage, B.D., 1982. Colonization of the large bowel by *Clostridium difficile* in healthy infants: quantitative study. *Infect. Immun.* 35, 895–899.

Starr, J.M., Martin, H., McCoubrey, J., Gibson, G., Poxton, I.R., 2003. Risk factors for *Clostridium difficile* colonisation and toxin production. *Age Ageing* 32, 657–660.

Stauffer, D.A., Karlin, A., 1994. Electrostatic potential of the acetylcholine binding sites in the nicotinic receptor probed by reactions of binding-site cysteines with charged methanethiosulfonates. *Biochemistry* 33, 6840–6849.

Stryer, L., Haugland, R.P., 1967. Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. U. S. A.* 58, 719–26.

Stubbs, S., Rupnik, M., Gibert, M., Brazier, J.S., Duerden, B., Popoff, M.R., 2000. Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of *Clostridium difficile*. *FEMS Microbiol. Lett.* 186, 307–312.

Sun, X., Wang, H., Zhang, Y., Chen, K., Davis, B., Feng, H., 2011. Mouse relapse model of *Clostridium difficile* infection. *Infect. Immun.* 79, 2856–2864.

Suree, N., Liew, C.K., Villareal, V.A., Thieu, W., Fadeev, E.A., Clemens, J.J., Jung, M.E., Clubb, R.T., 2009. The structure of the *Staphylococcus aureus* sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. *J. Biol. Chem.* 284, 24465–24477.

Tamayo, R., Pratt, J.T., Camilli, A., 2007. Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu. Rev. Microbiol.* 61, 131–48.

Tedesco, F.J., Barton, R.W., Alpers, D.H., 1974. Clindamycin-associated colitis. A prospective study. *Ann. Intern. Med.* 81, 429–433.

Teneberg, S., Lönnroth, I., Torres López, J.F., Galili, U., Halvarsson, M.O., Angström, J., Karlsson, K.A., 1996. Molecular mimicry in the recognition of glycosphingolipids by Gal alpha 3 Gal beta 4 GlcNAc beta-binding *Clostridium difficile* toxin A, human natural anti

alpha-galactosyl IgG and the monoclonal antibody Gal-13: characterization of a binding-active human *Glycobiology* 6, 599–609.

Thammavongsa, V., Kern, J.W., Missiakas, D.M., Schneewind, O., 2009. *Staphylococcus aureus* synthesizes adenosine to escape host immune responses. *J. Exp. Med.* 206, 2417–27.

Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F., Schneewind, O., 1999. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc. Natl. Acad. Sci. U. S. A.* 96, 12424–9.

Ton-That, H., Marraffini, L.A., Schneewind, O., 2004. Sortases and pilin elements involved in pilus assembly of Corynebacterium diphtheriae. *Mol. Microbiol.* 53, 251–261.

Ton-That, H., Mazmanian, S.K., Alksne, L., Schneewind, O., 2002. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *J. Biol. Chem.* 277, 7447–7452.

Ton-That, H., Mazmanian, S.K., Faull, K.F., Schneewind, O., 2000. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates. *J. Biol. Chem.* 275, 9876–81.

Tremblay, L.W., Xu, H., Blanchard, J.S., 2010. Structures of the Michaelis complex (1.2 Å) and the covalent acyl intermediate (2.0 Å) of cefamandole bound in the active sites of the Mycobacterium tuberculosis β-lactamase K73A and E166A mutants. *Biochemistry* 49, 9685–9687.

Tulli, L., Marchi, S., Petracca, R., Shaw, H.A., Fairweather, N.F., Scarselli, M., Soriani, M., Leuzzi, R., Vaccines, N., Srl, D., Bacteriology, M., 2013. CbpA: a novel surface exposed adhesin of *Clostridium difficile* targeting human collagen 1. *Cell Microbiol.* 15 1674-1687

Vandekerckhove, J., Schering, B., Bärmann, M., Aktories, K., 1988. Botulinum C2 toxin ADP-ribosylates cytoplasmic beta/gamma-actin in arginine 177. *J. Biol. Chem.* 263, 696–700.

Vollmer, W., Joris, B., Charlier, P., Foster, S., 2008. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol. Rev.* 32, 259–286.

Waligora, A.J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A., Karjalainen, T., 2001. Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. *Infect. Immun.* 69, 2144–2153.

Walk, S.T., Micic, D., Jain, R., Lo, E.S., Trivedi, I., Liu, E.W., Almassalha, L.M., Ewing, S.A., Ring, C., Galecki, A.T., Rogers, M.A.M., Washer, L., Newton, D.W., Malani, P.N., Young, V.B., Aronoff, D.M., 2012. *Clostridium difficile* ribotype does not predict severe infection. *Clin. Infect. Dis.* 55, 1661–1668.

Walker, K.J., Gilliland, S.S., Vance-Bryan, K., Moody, J.A., Larsson, A.J., Rotschafer, J.C., Guay, D.R., 1993. *Clostridium difficile* colonization in residents of long-term care facilities: prevalence and risk factors. *J. Am. Geriatr. Soc.* 41, 940–946.

Warny, M., Pepin, J., Fang, A., Killgore, G., Thompson, A., Brazier, J.S., Frost, E., McDonald, L.C., 2005. Toxin production by an emerging strain of *Clostridium difficile* associated with

outbreaks of severe disease in North America and Europe. *Lancet* 366, 1079–1084.

Warny, M., Vaerman, J.P., Avesani, V., Delmee, M., 1994. Human antibody response to *Clostridium difficile* toxin A in relation to clinical course of infection. *Infect. Immun.* 62, 384–389.

Weeks, C., Miller, R., S, M., 1999. Optimizing Shake-and-Bake for proteins. *Acta Crystallogr. D. Biol. Crystallogr.* 55, 492–500.

Wegner, A., Aktories, K., 1988. ADP-ribosylated actin caps the barbed ends of actin filaments. *J. Biol. Chem.* 263, 13739–13742.

Weiner, E.M., Robson, S., Marohn, M., Clubb, R.T., 2010. The Sortase A enzyme that attaches proteins to the cell wall of *Bacillus anthracis* contains an unusual active site architecture. *J. Biol. Chem.* 285, 23433–23443.

Weiner, M.P., Costa, G.L., 1994. Rapid PCR site-directed mutagenesis. *PCR Methods Appl.* 4, S131–S136.

Wenisch, C., Parschalk, B., Hasenhündl, M., Hirschl, A.M., Graninger, W., 1996. Comparison of vancomycin, teicoplanin, metronidazole, and fusidic acid for the treatment of *Clostridium difficile*-associated diarrhea. *Clin. Infect. Dis.* 22, 813–818.

Wilcox, M.H., Cunniffe, J.G., Trundle, C., Redpath, C., 1996. Financial burden of hospital-acquired *Clostridium difficile* infection. *J. Hosp. Infect.* 34, 23–30.

Wilcox, M.H., Fawley, W.N., Settle, C.D., Davidson, A., 1998. Recurrence of symptoms in *Clostridium difficile* infection--relapse or reinfection? *J. Hosp. Infect.* 38, 93–100.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 235–242.

Winter, G., Lobley, C.M.C., Prince, S.M., 2013. Decision making in xia2. *Acta Crystallogr. D. Biol. Crystallogr.* 69, 1260–1273.

Wren, B.W., 2011. Clostridium Gene - US 20120301428 A1. US 20120301428 A1.

Wright, A., Wait, R., Begum, S., Crossett, B., Nagy, J., Brown, K., Fairweather, N.F., 2005. Proteomic analysis of cell surface proteins from *Clostridium difficile*. *Proteomics* 5, 2443–2452.

Wright, J.S., Jin, R., Novick, R.P., 2005. Transient interference with staphylococcal quorum sensing blocks abscess formation. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1691–1696.

Xiao, Q., Jiang, X., Moore, K.J., Shao, Y., Pi, H., Dubail, I., Charbit, A., Newton, S.M., Klebba, P.E., 2011. Sortase independent and dependent systems for acquisition of haem and haemoglobin in Listeria monocytogenes. *Mol. Microbiol.* 80, 1581–1597.

Zar, F.A., Bakkanagari, S.R., Moorthi, K.M.L.S.T., Davis, M.B., 2007. A comparison of

vancomycin and metronidazole for the treatment of *Clostridium difficile*-associated diarrhea, stratified by disease severity. *Clin. Infect. Dis.* 45, 302–307.

Zhang, R., Wu, R., Joachimiak, G., Mazmanian, S.K., Missiakas, D.M., Gornicki, P., Schneewind, O., Joachimiak, A., 2004. Structures of sortase B from *Staphylococcus aureus* and *Bacillus anthracis* reveal catalytic amino acid triad in the active site. *Structure* 12, 1147–1156.

Zimmerly, S., Guo, H., Perlman, P.S., Lambowitz, A.M., 1995. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* 82, 545–554.

Zink, S.D., Burns, D.L., 2005. Importance of srtA and srtB for growth of *Bacillus anthracis* in macrophages. *Infect. Immun.* 73, 5222–5228.

Zong, Y., Bice, T.W., Ton-That, H., Schneewind, O., Narayana, S.V.L., 2004a. Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. *J. Biol. Chem.* 279, 31383–31389.

Zong, Y., Mazmanian, S.K., Schneewind, O., Narayana, S. V, 2004b. The structure of sortase B, a cysteine transpeptidase that tethers surface protein to the *Staphylococcus aureus* cell wall. *Structure* 12, 105–112.

Appendix A – Nucleotide and Amino Acid Sequences

A1.1 - Nucleotide Sequences

A1.1.1 - DNA Sequence of *srtB* insert of pEXP1-SrtB:

ATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACG
ACGATGACGATAAGGATCATCAAACAAGTTTGTACAAAAAAGCAGGCTCCGCGGCCGCCCCCTTCACCCACGACACCAA
AATCTCTAGCGAGCTGCAGAAAAAAGAGTACAAAAAAGAAGACCTGAGCAAAATCAACAGCGACTTCAAATTCTGGCTG
TCTGTGGAAAACACTAACATCAACTATCCAGTGGTACAGAGCAAAGATAACTCCTACTATCTGGACAAAGACTTCTACA
AAAAAGACTCTATCAGCGGTACTCTCTTTATGGATTACCGCAACAAATCTATCGACGACAAAAACATTATCATCTACGG
TCACAACATGAAAAACAAAACGATGTTCAACAACCTGAACAAATTCAAAGACGCAGACTTCTTCAAAAAAAACAACAAA
ATCAAAATCACGCTCAACGGTAAAGAGTTTCTGTACGACGTATTCTCTGCGTATATTGTCGAATCCGATTACGACTACC
TCAAAACCAACTTCAACAACGAATCTGACTATCAGAACTACATCAACGATATCACGAGCAAATCTCTCTACAAATCTCC
AATCAAAGTGAACTCCAACGACAAAATCGTTACGCTGTCTACTTGTACTTACGAATTCGATGACGCTCGTATGGTAATT
CACGGTCGTCTGATCTAAAAGGGTGGGCGCGCCGACCCAGCTTT

A1.1.2 - DNA Sequence of srtB C226A insert of pEXP1-SrtBC226A:

ATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACG
ACGATGACGATAAGGATCATCAAACAAGTTTGTACAAAAAAGCAGGCTCCGCGGCCGCCCCCTTCACCCACGACACCAA
AATCTCTAGCGAGCTGCAGAAAAAAGAGTACAAAAAAGAAGACCTGAGCAAAATCAACAGCGACTTCAAATTCTGGCTG
TCTGTGGAAAACACTAACATCAACTATCCAGTGGTACAGAGCAAAGATAACTCCTACTATCTGGACAAAGACTTCTACA
AAAAAGACTCTATCAGCGGTACTCTCTTTATGGATTACCGCAACAAATCTATCGACGACAAAAACATTATCATCTACGG
TCACAACATGAAAAACAAAACGATGTTCAACAACCTGAACAAATTCAAAGACGCAGACTTCTTCAAAAAAAACAACAAA
ATCAAAATCACGCTCAACGGTAAAGAGTTTCTGTACGACGTATTCTCTGCGTATATTGTCGAATCCGATTACGACTACC
TCAAAACCAACTTCAACAACGAATCTGACTATCAGAACTACATCAACGATATCACGAGCAAATCTCTCTACAAATCTCC
AATCAAAGTGAACTCCAACGACAAAATCGTTACGCTGTCTACTGCTACTTACGAATTCGATGACGCTCGTATGGTAATT
CACGGTCGTCTGATCTAAAAGGGTGGGCGCGCCGACCCAGCTTT

A.1.1.3 – DNA sequence of CD0386 insert of pJ414 (C-terminal Tag)

AGGAGGTAAAACATATGAAGCTTGAGAGCAAACAATACTGGACGGAGAGCGCGGAAAGAGTGGGTATTATCGAAAAGGT
GATGAATGACGGCAGCATTGGTAGCACCTTTAACGAGGGCTATATGAAAGTCGAGGGTGAAACGGCATATTGCATCGAC
ATCAACACCGATTTCAAAAATGGCTATAAGACCCGTACCGATGCGTCCAGCCGCATGAGCGCGGATCAAATTAGCGACG
TAGCGCTGAGCCTGGAGTACGTGAAGCAGTATGGCGAGGCCCATAAGGAACTGAACTATAAACAAGTGTACCTGCTGGA
ACAATGTGTCGTCTGGCAGCGTTTGAGCGTCCACCTGGGTTGGCAATGTGACAATGTCCGCGCCTCTTACGACGAGATC
CCGAAAGCTACTCAGGACGAAGTGTTCGCAGGTGCAAAGGCGTTTGTCAAAGAGAACAAAGGTCGTTATGAATGCGGTG
GTTACATCTACTCCGGCGAGGGCCAGGAGCTGGGTCAGTTCTGGGCGAAGCTGAATGTGGGTAACGCGAAACTGCAGAA
AACCAGCAGCAATGCTTCTATTACCGATGGTAACGGTAACTACTCCATCGCGGGTGCGACCTACGGTGTGTTTGCGGAC
AAGGACTGCACGAAGCAACTGGCCACCTTGACCACCAACGAGAATGGTAATACGGACGTCGTTGAGGTCACGACTGGTA
CGGTTTACATCCGTGAATTGAGCGCACCGGCTGGCTACAAAGTCGACAAAACCGTGTATAGCCTGAAGGTTGAAGCGGG
CAAAACCGCAACTCTGAAAGTTAGCGACACCCCGAAAGTTACCGATACCCTGATTGAGCTGTTCAAAATTGATATGGAA
ACCCAAAAAGACAACCCGCAGGGCAATGCCAGCCTGGCCGGTGCCGAGTTTACCTGGAAGTATTATGCGGGTTTCTATA
ACAAGGAAAACCTGCCAGCGGAGGCGACGCGCACGTGGGTTACCAAAACGATCGCAGAGACTGATAGCGATGGTACGAC
CCACTACATTACCAAGCTGGCGGACGCTTATAAGGTTTCTGGTGATTCCTTCTATATGCAGGACGGTAAGGCAGTGCTG
CCGCTGGGTACTTTGACTGTTGAAGAAACGAAGGCCCCGAATGGTTATCTGCTGGATGGTGCGTACATGCAAGCCGGTG
ATAAAAGCGAGCAAATCAAGGGCCTGTATGTCACTCAGATCACCGAGGATGGTGATCTGGCGGTGCTGTCCGGTAGCAA
TCAGTTTAGCGTGAGCGATAAAGTCATTCGTGGTGGTGTCAAGATTCAGAAGCGCGATCTGGAAACCGGTGATACCAAG
CCTCAAGGTAGCGCGACGCTGAAGGACACGGCTTTTGACATCATCAGCTTGAACGATAACGTCGTGCTGGTGGAGGGCA
AACTGTACAAGAAAAACGAAGTCGTTAAAACGATCCACACCGATATCGAGGGTGTCGCCAGCACGTCGGCCGATCTGCT
GCCGTACGGTAAGTTTCGCATTGTGGAATCGGAAGCTCCTAATGGCTATCTGACCGATGGCGCAAAACCGATCGATTTT
GCAATCACCGAGAATGGCAAGATTGTTGATCTGACGGACGAGGCGCACAGCATTTACAACCAGATTAAACGTGGCGATA
TCGAAGGTGTGAAGATCGGTGCGGGCACGCACAAACGTCTGGCTGACGTTCCGTTTCGTATCACCAGCAAGACGACGGG
CGAGAGCCACGTTGTTGTCACCGACGACAATGGTCAGTTCAGCACCAGCGCGGACTGGGCGAGCCATAAGCATAACACG
AATGCAGGTAAGACCAGCGAGGACGGCGTTTGGTTTGGTACGTCGGAACCGGATGATTCGAAGGGTGCCCTGCCGTATG
ATACCTACATTATCGAAGAGCTGCGTAGCGACAGCAATAAGGGCTTCGAGCTGATCCCACCTTTCGAGATCGTGGTGTC
GCGTAACAATCTGGTTATTGACCTGGGCACCCTGACCGACGAGTACGAGAAAGAGATCAGCATTCATACGACCGCAAAC
AGCAAGGATGGCGAAAAGACCATCTTGGCGGGTAAAGAGGTTACGATTGTGGATACCGTTAAACTGGACGGCCTGACCA
AAGGCACCAAATACCAGCTGAAAGGTTGGCAAATGCTGAAGGAAGAAAACGCGGAGCTGATTATTGATGGTAAGCGTGT

CGAGAACGATTACACGTTTATCGCGGACGATGAAGCAATGAAAGTGGAAATTGCGTACACGTTCAACGCCTCTGAACTG
GGCGGTAAAAACTTGGTCACCTTTGAGGAACTGTACGACCTGAGCAACCCGGACGAGCCGGTTAAGGTCGCCGAGCATA
AAGACATTGCAGACGATGGTCAGACCGTGTTGATTACCGAACGCATTATCAAAATTCACACCACCGCGACCGACAAGGA
CGGCAATAAAGAGTTGGAAGCCGGTAAGGAAGTCACGATTATCGATACCATTACCCTGGAAGGTCTGGAAGTTGGCACG
CAATACAAACTGGTTGGCTGGCAGATGCTGAAAGAGGAGAATGCAGAGCTGCTGATCAACGGCAAGCGTGTGGAGTCCG
ACTATACGTTTACCGCTGACAGCGAAACCATGAAGGTGGAGGTTGCATTCACCTTCGACGCGACCAGCCTGGATGGCAA
ACAACTGGTTACGTTCGAGGAATTGTACGATCTGAGCAATCCGGACGAACCGAAGAAAGTGACCGAACACAAGGACATC
GAGGACGAGGGTCAGACGATTACCTTCAAGGAAAAACCGGAGGTCCCGGAAGAACCGGAAACCCCGCAAACCCCGGAAA
AGCCGAGCCGTCCGAGCGATTCCCCGAAAACGGGTGACAATACGAACCTGTATGGCCTGTTGGCACTGTTGCTGACGAG
CGGTGCTGGTCTGGCGGGTATCTTCTTCTACAAACGTCGCAAAATGAAGAAAAGCGGTACCCTGGTTCCGCGTGGTTCT
CACCACCACCACCACCACTAACTCGAG

A1.2 Amino Acid Sequences

A1.2.1 - Amino Acid Sequence of SrtB as expressed from pEXP1:

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDHQTSLYKKAGSAAAPFTHDTKISSELQKKEYKKEDLSKINSDFKFWL
SVENTNINYPVVQSKDNSYYLDKDFYKKDSISGTLFMDYRNKSIDDKNIIIYGHNMKNKTMFNNLNKFKDADFFKKNNK
IKITLNGKEFLYDVFSAYIVESDYDYLKTNFNNESDYQNYINDITSKSLYKSPIKVNSNDKIVTLSTCTYEFDDARMVI
HGRLI

Theoretical Mass = 28307.6

A1.2.2 - Amino Acid Sequence of SrtB C226A as expressed from pEXP1:

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDHQTSLYKKAGSAAAPFTHDTKISSELQKKEYKKEDLSKINSDFKFWL
SVENTNINYPVVQSKDNSYYLDKDFYKKDSISGTLFMDYRNKSIDDKNIIIYGHNMKNKTMFNNLNKFKDADFFKKNNK
IKITLNGKEFLYDVFSAYIVESDYDYLKTNFNNESDYQNYINDITSKSLYKSPIKVNSNDKIVTLSTATYEFDDARMVI
HGRLI

Theoretical Mass = 28275.5

A1.2.3 - CD0386 as expressed from pJ414 (C-terminal tag):

MKLESKQYWTESAERVGIIEKVMNDGSIGSTFNEGYMKVEGETAYCIDINTDFKNGYKTRTDASSRMSADQISDVALSL
EYVKQYGEAHKELNYKQVYLLEQCVVWQRLSVHLGWQCDNVRASYDEIPKATQDEVFAGAKAFVKENKGRYECGGYIYS
GEGQELGQFWAKLNVGNAKLQKTSSNASITDGNGNYSIAGATYGVFADKDCTKQLATLTTNENGNTDVVEVTTGTVYIR
ELSAPAGYKVDKTVYSLKVEAGKTATLKVSDTPKVTDTLIELFKIDMETQKDNPQGNASLAGAEFTWKYYAGFYNKENL
PAEATRTWVTKTIAETDSDGTTHYITKLADAYKVSGDSFYMQDGKAVLPLGTLTVEETKAPNGYLLDGAYMQAGDKSEQ
IKGLYVTQITEDGDLAVLSGSNQFSVSDKVIRGGVKIQKRDLETGDTKPQGSATLKDTAFDIISLNDNVVLVEGKLYKK
NEVVKTIHTDIEGVASTSADLLPYGKFRIVESEAPNGYLTDGAKPIDFAITENGKIVDLTDEAHSIYNQIKRGDIEGVK
IGAGTHKRLADVPFRITSKTTGESHVVVTDDNGQFSTSADWASHKHNTNAGKTSEDGVWFGTSEPDDSKGALPYDTYII
EELRSDSNKGFELIPPFEIVVSRNNLVIDLGTLTDEYEKEISIHTTANSKDGEKTILAGKEVTIVDTVKLDGLTKGTKY
QLKGWQMLKEENAELIIDGKRVENDYTFIADDEAMKVEIAYTFNASELGGKNLVTFEELYDLSNPDEPVKVAEHKDIAD
DGQTVLITERIIKIHTTATDKDGNKELEAGKEVTIIDTITLEGLEVGTQYKLVGWQMLKEENAELLINGKRVESDYTFT
ADSETMKVEVAFTFDATSLDGKQLVTFEELYDLSNPDEPKKVTEHKDIEDEGQTITFKEKPEVPEEPETPQTPEKPSRP
SDSPKTGDNTNLYGLLALLLTSGAGLAGIFFYKRRKMKKSGTLVPRGSHHHHHH

Theoretical Mass = 110589.3
Theoretical Mass following thrombin cleavage = 109622.3

A1.2.4 - CD0386 as expressed from pTAC-MAT1 (N-terminal tag):

MAHNHRKHMKLESKQYWTESAERVGIIEKVMNDGSIGSTFNEGYMKVEGETAYCIDINTDFKNGYKTRTDASSRMSADQ
ISDVALSLEYVKQYGEAHKELNYKQVYLLEQCVVWQRLSVHLGWQCDNVRASYDEIPKATQDEVFAGAKAFVKENKGRY
ECGGYIYSGEGQELGQFWAKLNVGNAKLQKTSSNASITDGNGNYSIAGATYGVFADKDCTKQLATLTTNENGNTDVVEV
TTGTVYIRELSAPAGYKVDKTVYSLKVEAGKTATLKVSDTPKVTDTLIELFKIDMETQKDNPQGNASLAGAEFTWKYYA

GFYNKENLPAEATRTWVTKTIAETDSDGTTHYITKLADAYKVSGDSFYMQDGKAVLPLGTLTVEETKAPNGYLLDGAYM
QAGDKSEQIKGLYVTQITEDGDLAVLSGSNQFSVSDKVIRGGVKIQKRDLETGDTKPQGSATLKDTAFDIISLNDNVVL
VEGKLYKKNEVVKTIHTDIEGVASTSADLLPYGKFRIVESEAPNGYLTDGAKPIDFAITENGKIVDLTDEAHSIYNQIK
RGDIEGVKIGAGTHKRLADVPFRITSKTTGESHVVVTDDNGQFSTSADWASHKHNTNAGKTSEDGVWFGTSEPDDSKGA
LPYDTYIIEELRSDSNKGFELIPPFEIVVSRNNLVIDLGTLTDEYEKEISIHTTANSKDGEKTILAGKEVTIVDTVKLD
GLTKGTKYQLKGWQMLKEENAELIIDGKRVENDYTFIADDEAMKVEIAYTFNASELGGKNLVTFEELYDLSNPDEPVKV
AEHKDIADDGQTVLITERIIKIHTTATDKDGNKELEAGKEVTIIDTITLEGLEVGTQYKLVGWQMLKEENAELLINGKR
VESDYTFTADSETMKVEVAFTFDATSLDGKQLVTFEELYDLSNPDEPKKVTEHKDIEDEGQTITFKEKPEVPEEPETPQ
TPEKPSRPSDSPKTGDNTNLYGLLALLLTSGAGLAGIFFYKRRKMKKS

Theoretical Mass = 110010.8

A1.2.5 - CD0386N as expressed from pET28a:

HHHHHHSSGLVPRGSHMMKLESKQYWTESAERVGIIEKVMNDGSIGSTFNEGYMKVEGETAYCIDINTDFKNGYKTRTD
ASSRMSADQISDVALSLEYVKQYGEAHKELNYKQVYLLEQCVVWQRLSVHLGWQCDNVRASYDEIPKATQDEVFAGAKA
FVKENKGRYECGGYIYSGEGQELGQFWAKLNVGNAKLQKTSSNASITDGNGNYSIAGATYGVFADKDCTKQLATLTTNE
NGNTDVVEVTTGTVYIRELSAPAGYKVDKTVYSLKVEAGKTATLKVSDTPKVTDTLIELFKIDMETQKDNPQGNASLAG
AEFTWKYYAGFYNKENLPAEATRTWVTKTIAETDSDGTTHYITKLADAYKVSGDSFYMQDGKAVLPLGTLTVEETKAPN
GYLLDGAYMQAGDKSEQIKGLYVTQITEDGDLAVLSGSNQFSVSDKVIRGGVKIQKRDLETGDTKPQGSATLKDTAFDI
ISLNDNVVLVEGKLYKKNEVVKTIHTDIEGVASTSADLLPYGKFRIVESEAPNGYLTDGAKPIDFAITENGKIVDLTDE
AHSIYNQIKRGDIEGVK

Theoretical mass = 62630.9
Theoretical mass following thrombin cleavage = 61111